Foundational approaches to Celtic linguistics

Edited by

Andrew Carnie
Diane Ohala
Dee Hunter
Samantha Prins
Michael Hammond
Luis Irizarry

Current Issues in Celtic Linguistics 1

language science press

Current Issues in Celtic Linguistics

Editors:
Andrew Carnie, University of Arizona
Michael Hammond, University of Arizona
Diane Ohala, University of Arizona

In this series:

1. Carnie, Andrew, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond & Luis Irizarry (eds.). Foundational approaches to Celtic linguistics.

ISSN (print):
ISSN (electronic):

Foundational approaches to Celtic linguistics

Edited by

Andrew Carnie
Diane Ohala
Dee Hunter
Samantha Prins
Michael Hammond
Luis Irizarry

language
science
press

Andrew Carnie, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond & Luis Irizarry (eds.). 2025. *Foundational approaches to Celtic linguistics* (Current Issues in Celtic Linguistics 1). Berlin: Language Science Press.

This title can be downloaded at:
http://langsci-press.org/catalog/book/483
© 2025, the authors
Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):
http://creativecommons.org/licenses/by/4.0/
ISBN: 978-3-96110-515-1 (Digital)
978-3-98554-144-7 (Hardcover)

ISSN (print):
ISSN (electronic):
DOI: 10.5281/zenodo.15532992
Source code available from www.github.com/langsci/483
Errata: paperhive.org/documents/remote?type=langsci&id=483

Cover and concept of design: Ulrike Harbort
Typesetting: Andrew Carnie, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond, Luis Irizarry
Proofreading: David Carrasco Coquillat, Dmitry Nikolaev, Elen Le Foll, Elisa Roma, Elliott Pearl, Eva Schultze Berndt, Jean Nitzke, Jeroen van de Weijer, Katja Politt, Lachlan Mackenzie, Mary Ann Walter, Matthew Korte, Nicoletta Romeo, Rainer Schulze, Rebecca Madlener, Sebastian Nordhoff, Silvie Strauß
Fonts: Libertinus, Arimo, DejaVu Sans Mono
Typesetting software: XeLaTeX

Language Science Press
Scharnweberstraße 10
10247 Berlin, Germany
http://langsci-press.org
support@langsci-press.org

Storage and cataloguing done by FU Berlin

FREIE
UNIVERSITÄT
BERLIN

Contents

Preface
Andrew Carnie, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond & Luis Irizarry **v**

Acknowledgments **ix**

I Syntax

1 Polarity sensitivity and fragments in Irish
James McCloskey **3**

2 When linearization triggers embedded V2: Evidence from Breton
Melanie Jouitteau **63**

3 Deriving VSO in Scottish Gaelic
Gillian Catriona Ramchand **97**

4 Reassessing Oehrle effects: Evidence from Scottish Gaelic
Gary Thoms **141**

5 Some complexities and idiosyncrasies of Welsh consonantal mutation
Maggie Tallerman **183**

II Phonology and phonetics

6 Phonology of the Celtic languages
Michael Hammond **211**

7 Hebrides English in the 1980s and now
Ian Clayton & Cynthia Shuken **247**

8 Gestural timing and contrast: An Irish case study
Ryan Bennett, Jaye Padgett, Grant McGuire, Máire Ní Chiosáin & Jennifer Bellik **283**

III Language change, historical linguistics, and grammaticalization

9 Comparing the syntactic complexity of Gaeltacht and urban Irish-Language broadcasters
Brian Ó Broin **345**

10 Bunadas: A network database of cognate words, with emphasis on Celtic
Caoimhín P. Ó Donnaíle **369**

11 Laryngeal contrast in the Celtic languages: Variation, typology, and history
Pavel Iosad **383**

12 Syntactic reconstruction in Celtic
Marieke Meelen **417**

IV Sociolinguistics and language revitalization

13 *Guth Thormoid*: The "Island Voice" of Norman Maclean
Gordon Wells **471**

14 White supremacists' weaponization of Celtic heritage languages and language revitalization
John Powell **495**

15 France's war on Breton diacritics: An incomprehensible obstinacy
Stefan Moal **527**

Index **554**

´

Preface

Andrew Carnie, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond & Luis Irizarry[a]
[a]University of Arizona

The Celtic Linguistics Group was founded at the University of Arizona in Spring 2006 and has had over 25 affiliated faculty, post-docs, graduate students, and undergraduates, as well as strong community partnerships with linguists and Celtic language speakers in Ireland, Scotland, Wales and Brittany. As such, it is a broad-based and interdisciplinary collective that has emphasized the importance of community-based linguistics from its inception. The group has sponsored nine US National Science Research Grants and Supplements, which have included work in instrumental phonetics, experimental and theoretical phonology, morphology, syntax, psycholinguistics, field and documentary linguistics, and language revitalization. It has also hosted two major conferences on the Celtic languages. Work completed by group members has resulted in several doctoral dissertations, numerous books and monographs, and countless articles and conference presentations.

Like many others, much of our research was dragged to a screeching halt with the onset of the pandemic in 2020. As a consequence, we decided that the best course of action to keep our community vibrant was to host a series of 31 virtual talks, bringing in experts in Celtic linguistics from a wide range of areas. This talk series came to be known as Foundational Approaches to Celtic Linguistics (FACL). Many of these talks can still be seen at https://www.youtube.com/@foundationalapproachestoce1421. As life returned to normal, we invited the speakers of these talks to contribute to this volume. In a few cases, we allowed some of our speakers to contribute different papers than they had presented at the time, if they had published their original FACL talks elsewhere. The result is a collection of 15 influential articles by some of the most prominent researchers in the field of Celtic linguistics.

Andrew Carnie, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond & Luis Irizarry. 2025. Preface. In Andrew Carnie, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond & Luis Irizarry (eds.), *Foundational approaches to Celtic linguistics*, v–viii. Berlin: Language Science Press.
DOI: 10.5281/zenodo.15654981

The first five chapters of the book are devoted to syntax. Jim McCloskey opens with an empirically rich and theoretically important work on the licensing of Negative Polarity items (NPIs) in Irish (*Polarity sensitivity and fragments in Irish*). Irish has the typologically unique property that NPIs can serve as fragment answers to questions. McCloskey shows that this property falls out from the fact that Irish allows a narrative fronting operation, which in turn gives rise to a sluice-like construction that looks like a fragment. Melanie Jouitteau's contribution (*When linearization triggers embedded V2: Evidence from Breton*) also deals with stylistic fronting operations and their interaction with Tense-second (T2) orders in embedded contexts, this time in Breton. She compares word order effects in different dialects and shows that the relevant data emerge from the fact that T2 linearization is a post-syntactic effect dependent upon prosody. The derivation of Verb Initial order is the focus of Gillian Ramchand's *Deriving VSO in Scottish Gaelic*. On the basis of data from pronoun and adverbial placement, she proposes a novel direct linearization algorithm that is sensitive to prosodic patterning. Gary Thoms' work (*Reassessing Oehrle effects: Evidence from Scottish Gaelic*) uses the fact that Gaelic completely lacks Oehrle effects to show that this cannot be due to semantic incompatibility. He argues that their absence in Gaelic is due to the height of subjects and their interaction with shifted objects. Maggie Tallerman's *Some complexities and idiosyncrasies of Welsh consonant mutation* ably transitions us between the syntax and phonology sections of the book. She champions the idea that initial consonant mutation cannot be solely characterized by triggers and explores mutations without triggers, with unrealized triggers, and with exceptions in loan words.

The second part of the book highlights critical issues in the phonetics and phonology of the Celtic languages. *Phonology of the Celtic languages* by Michael Hammond provides a comprehensive overview of the phonology of each of the Modern Celtic languages, comparing and synthesizing their similarities and differences. Ian Clayton and Cynthia Shuken's *Hebrides English in the 1980s and now* has a narrower focus on a variety of English that is heavily influenced by contact with Scottish Gaelic. Their chapter investigates the use of preaspiration, devoicing, and glottalization in Hebrides English and explores how these features have changed in the past 45 years. Of notable interest are the status of these features over time as either waxing or waning in prestige depending on the speaker population. The final contribution in this section, *Gestural timing and contrast: An Irish case study* by Ryan Bennet, Jaye Padgett, Grant McGuire, Máire Ní Chiosáin, and Jennifer Bellik, examines in great detail the extensive secondary articulation contrasts of palatalization and velarization found in Irish stop consonants. Specifically, the chapter analyzes the timing of the gestural properties of these

consonants as a function of syllable position and place of articulation and ably demonstrates that gestural timing is optimized to both maximize the contrasts and make them recoverable.

Section 3 is focused on issues in language change, diachronic linguistics and grammaticalization. Brian Ó Broin's work, *Comparing the syntactic complexity of Gaeltacht and urban Irish-Language broadcasters*, uses the Rosenberg and Abbeduto complexity scale to examine the differences between traditional rural Gaeltacht communities and those of fluent speakers in urban areas. He shows that urban broadcasters are simplifying the grammar by moving from hypotactic to paratactic constructions. Caoimhín Ó Donnaíle's chapter, *Bunadas: A network database of cognate words with emphasis on Celtic*, focuses on the Bunadas tool that he has developed for investigating etymological links between words in the different Celtic languages. As such, Ó Donnaíle's chapter provides a uniquely practical contribution to this volume. In *Laryngeal contrast in the Celtic languages: Variation, typology and history*, Pavel Iosad examines distinctions among the stop series of consonants in the Celtic languages over time to evaluate the likeliest onset of the present day aspirating system. The chapter provides an extensive synchronic and diachronic analysis of the relevant data that ultimately argues for an initial voicing-based system among the stop series, which gives way to the aspirating system at a point of key phonological change. Marieke Meelen's *Syntactic Reconstruction in Celtic* also focuses on reconstruction, but here takes on the notoriously tricky topic of syntactic reconstruction. She argues that recent generative approaches, such as the Parametric Comparison Method, shed light on the emergence of verb initial orders.

The final section of the book investigates sociolinguistic aspects of Celtic linguistics, with chapters on both language documentation efforts and critical issues in language use. Gordon Wells' chapter (*Guth Thormoid: The "Island voice" of Norman MacLean*) provides a case study that highlights the exceptional value of working closely with an experienced native speaker to not only provide linguistic data but to document vanishing cultures and values also impacted with language loss. John Powell's chapter (*White supremacists' weaponization of Celtic heritage languages*) provides an unsettling but necessary look into the minds of white supremacists who co-opt Celtic languages, language revitalization, and cultural documentation to further their own agendas. As such, the chapter is a clarion call to academic researchers about the need to pay careful attention to the gross misuse of the languages we treasure and the work we do. Stefan Moal's *France's war on Breton diacritics: An incomprehensible obstinacy* closes the book with a thorough history of the politicization of diacritic use in minority language names in France. Moal's work underscores how the systematic suppression of

Breton and other minority languages can result from unnecessary but deliberate governmental regulation of key orthographic features.

As can be seen, the chapters in this volume vary greatly in terms of their theoretical orientation, methodologies, and scope. This aptly reflects our group's continued commitment to a multidimensional and interdisciplinary perspective on the study of the Celtic languages as a means of providing the greatest understanding, support, and impact. Thus, some chapters offer summaries or reviews in particular domains of Celtic linguistics, whereas others are technical descriptions of specific methodologies and developmental projects. Still others are in-depth reports on original research or analyses of critical issues in the sub-fields of Celtic linguistics. This variety in approach showcases the diversity of the field and mirrors that of the vibrant research being conducted on the Celtic languages.

Acknowledgments

Pulling together an edited volume is a surprisingly complicated task that requires significant coordination among people who probably don't receive enough credit. They all have our most profound thanks.

We are extremely grateful to the entire team at Language Science Press. Sebastian Nordhoff and Felix Kopecky have been incredibly responsive in assisting us in getting this volume to production. They have never hesitated to answer our stupidest questions about LaTeX and have helped us debug this document every time we've introduced a crash-inducing formatting command. Our thanks to the devoted team of community proof-readers (David Carrasco Coquillat, Dmitry Nikolaev, Elen Le Foll, Elisa Roma, Elliott Pearl, Eva Schultze Berndt, Jean Nitzke, Jeroen van de Weijer, Katja Politt, Lachlan Mackenzie, Mary Ann Walter, Matthew Korte, Nicoletta Romeo, Rainer Schulze, Rebecca Madlener, Sebastian Nordhoff, and Silvie Strauß) for catching our mistakes and to the production team at the press. Any remaining errors are our own.

We offer a huge thank you to all of our authors, all of whom were exceptionally amenable to our requests and patient as well - and without whom this volume would not have been possible. We would also like to thank all the other presenters at the Foundational Approaches to Celtic Linguistics series who didn't contribute to this volume, but who also helped make the series so successful: David Adger, Joe Eska, S.J. Hannahs, Holly Kennard, Will Lamb, Christopher Lewin, Kerry McCullough, Bernadette O'ROurke, Claire Nance, Kevin Rottet, Kevin Scannell, Sylvia Schreiner, Cassie Smith-Christmas, Thomas Stewart, John Walsh, Peredur Webb-Davies, and David Willis. We'd also like to thank the audiences who attended these talks and provided such a rich environment for discussion and debate.

We owe particular gratitude to Kerry McCullough, who was a critical force in organizing the speaker series. She helped select and schedule speakers, organized students to introduce them, made our logo, and handled our advertising. Other students who also assisted us include Jesús González Franco, Andrea Fulgham, Brandon Gauthier, Luis Irizarry, Jennifer Medina, Alex Tubens Blanco, Colleen Patton, Sam Prins, John Powell, Dee Hunter, and Andrew Zupon.

High quality chapters require high quality reviewers, we relied on the following people to give detailed feedback and reviews on the papers. Thank you to: Ryan Bennett, Sonya Bird, Andrew Carnie, Ian Clayton, Joe Eska, Michael Hammond, S. J. Hannahs, Heidi Harley, Dee Hunter, Luis Irizarry, Jim McCloskey, Diane Ohala, John Powell, Samantha Prins, Natasha Warner, David Willis, and Andrew Zupon.

Go raibh maith agaibh, tapadh leibh, diolch yn fawr, trugarez deoc'ch, gura mie orroo, meur ras, merci and thank you!

— Andrew, Dee, Diane, Luis, Mike and Sam Tucson, March 2025

Part I

Syntax

Chapter 1

Polarity sensitivity and fragments in Irish

James McCloskey
University of California, Santa Cruz

In Irish, negative polarity items may serve, isolated, as fragment answers, something which seems initially to be both theoretically and typologically anomalous. This chapter explores some theoretical consequences of this observation. It begins by arguing that the apparent anomaly is real; that is, that the relevant fragments are in fact negative polarity items. It then argues that the anomaly can be resolved. The possibility of fragment negative polarity items is parasitic on an independent property of the language, namely that it possesses a movement operation (so-called Narrative Fronting) by which a narrow scope indefinite may be raised to a position to the left of the verb in a finite negative clause. It is argued that this option feeds a sluicing-like process of clausal ellipsis, and that this operation may strand licensed negative polarity items. In the context of 'move and delete' analyses of fragments, as developed especially by Merchant 2004, the apparently anomalous possibility is routine. A by-product of the discussion is a more complete mapping of the landscape of polarity sensitive items in Irish than has so far been provided.

1 Background and goals

The English question-answer pair in (1) is ill-formed:

(1) A: What did the priest say? B: *Anything.

Its apparent Irish counterpart in (2), however, is routine and well-formed:

(2) A: Cad a dúirt an sagart? B: Aon rud.
what C say.PAST the priest any thing
A: 'What did the priest say?' B: 'Anything'. CHD 57

James McCloskey. 2025. Polarity sensitivity and fragments in Irish. In Andrew Carnie, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond & Luis Irizarry (eds.), *Foundational approaches to Celtic linguistics*, 3–61. Berlin: Language Science Press.
DOI: 10.5281/zenodo.15654857

To express what Irish expresses by way of (2), English must use (3), exploiting the presence in its lexicon of the inherently negative expression *nothing*, used here as a fragment answer:

(3) A: What did the priest say? B: Nothing.

This chapter is concerned with the theoretical issues raised by the contrasts in (2–3). Those issues are important in part because the Irish possibility shown in (2) seems to represent a theoretical and typological anomaly. Negative polarity items are not supposed to be able to function, in isolation, as fragment answers.[1]

The first goal of this chapter is to establish that this interesting and unexpected possibility does in fact exist; the answer in (2) really does consist solely (in its overt form) of a negative polarity item. Its second goal is to develop a way of understanding that possibility which resolves the apparent anomaly. Its third goal is to consider some larger implications which flow from that account for the general theory of subsentential fragments and for the theory of ellipsis.

A by-product of the discussion will be a more complete map of the landscape of polarity sensitive items in Irish than has so far been available. Most discussions of polarity sensitivity currently available deal with languages that have among their lexical resources expressions which are inherently negative, such as English *nobody*, *nothing* or *never*. I will argue here, though, that Acquaviva (1996) was right to claim that Irish has no such expressions. We are thus presented with an opportunity to explore what patterns emerge in their absence.

Two sources of data form the basis for that exploration here: work with six native speaker consultants over a period of several years, along with a collection of 1800 naturally occurring examples of polarity sensitive items of various kinds and in various contexts of use.[2]

Some terminological preliminaries to begin with. I will use the term 'polarity sensitive expression' (PSE) for such elements as *aon rud* in 2 or *never* in English. These are expressions whose interpretation and well-formedness depend on the polarity of the larger environment in which they find themselves. The set of PSEs is the superset which includes as subsets both negative polarity items (NPIs) and 'inherently negative expressions' (INEs), which is the term I will use here for the elements called variously '*n*-words' or 'negative indefinites' or 'negative

[1]The star on English (1) is well-deserved, but there are conditions under which NPIs may appear as fragment answers in English. This issue is taken up in Section 4.4 below.

[2]Examples from this corpus are indicated by a tag consisting of an abbreviation of the title (in the case of published texts) followed by a number indicating the page from which the example was extracted. The abbreviations used in these tags are explained in Appendix Two.

quantifiers' in other discussions (English *nothing* or French *personne*). In some languages and varieties, such expressions may stand alone to express sentential negation; in others, they must or may enter into negative concord dependencies to serve that function.

2 The landscape of polarity sensitivity in Irish

In presenting the contrast between Irish (2) and English (1), I glossed the Irish *aon rud* as English 'any thing'. If I had instead chosen to use 'no thing' as the gloss, I could have created the impression that there was no anomaly to be concerned about, since many languages allow inherently negative expressions to function as fragment answers. The glossing decision, then, mirrors an important theoretical issue in this case. Consider the pair of Irish examples in (4)[3]:

(4) a. Ní-or iarr duine ar bith orm é.
C.NEG-PAST ask.PAST person any on.me it
'Nobody asked me for it.'

b. * D' iarr duine ar bith orm é.
PAST ask.PAST person any on.me it

Since (4b) is impossible, we know that *duine ar bith* cannot express negation (and is, in fact, ill-formed) when not within an appropriate licensing environment. In (4a) it appears in just such an environment – that determined by the clause-initial sentential negation. It is therefore well-formed and the meaning it expresses is that of the English translation. *Duine ar bith* in (4a) must then be either a negative polarity item or an INE in a concord dependency with the clause-initial negation (e.g., Labov 1972; Laka 1990; Haegeman & Zanuttini 1991; Ladusaw 1992; Zanuttini 1997; Penka & Zeijlstra 2010; Zeijlstra (2004; 2008; 2016); Giannakidou & Zeijlstra 2017; Deal 2022). But, given the possibility of (2), to maintain consistency with the received wisdom concerning fragment answers, we must assume that *aon rud* (and many similar expressions we will encounter

[3]The glossing conventions used are, for the most part, familiar and transparent but a few notes are in order. Because the inflected prepositions are not a matter of concern here, they are glossed for readability using English pronouns. So Irish *liom* is glossed as 'with.me'. The 'direct relative' complementizer is glossed as C.WH and the marker of polar interrogatives as C.Q. All other complementizers, including the conditional, are simply glossed as C. Past and conditional forms of the complementizers are segmentalized, so that, for instance, *gur* is spelled *gu-r* and glossed as C-PAST. 'Autonomous' forms of verbs are glossed as IMPERS and imperative forms as IMPV. The particle *a*, which precedes verbal nouns, is glossed as VCE (suggesting 'voice') and verbal nouns themselves are glossed with .VN. The progressive particle *ag* is glossed as PROG.

shortly) are not NPIs but are rather INEs. However, Acquaviva (1996) has argued that Irish (like Māori[4]) possesses no INEs. If he is right, it follows *a fortiori* that Irish cannot be a negative concord language. And it follows in turn that *duine ar bith* in (4) and *aon rud* in the fragment answer of (2) must be NPIs and that the apparent conflict with typological and theoretical expectation is real.

It is important, then, to determine whether the pattern in (4) reflects a negative concord system or an NPI-centered system, and the first business of the present chapter will be to address that question. Making such a distinction, though, is not as straightforward as it once seemed to be; the empirical and theoretical landscapes seem more intricate than they once did (see Laka 1993 and, especially, Herburger 2001). I will argue, though, that Acquaviva was correct about the absence of INEs in Irish, that all of its polarity sensitive expressions are in fact NPIs, and that the theoretical questions raised by (2) are therefore real.

Two convictions drive the discussion. The first is that the distribution of NPIs is determined by fundamentally semantic and pragmatic relations (Fauconnier (1975; 1979); Ladusaw 1979; Heim 1984; Kadmon & Landman 1993; Krifka 1995; Zwarts 1996; van der Wouden (1994; 1997); Giannakidou 1998; Lahiri 1998; von Fintel 1999; Hoeksema 2000; Gajewski (2005; 2011); Guerzoni & Sharvit 2007; Homer 2011: Chap. 2; Chierchia 2013; Crnič 2014; Gajewski & Hsieh 2014; Barker 2018; Crnič 2019; Homer 2021; Jeong & Roelefsen 2023). The second is that INEs, though they may well be strong NPIs in their semantics, are licensed in negative concord dependencies by featural interactions in the syntax, subject to characteristically syntactic requirements of locality and relative prominence. This is why analyses in terms of the operation AGREE have been so influential and seemed so persuasive in recent years (see Zeijlstra 2008 and recent exchanges among Zeijlstra 2012, Preminger & Polinsky 2015, Bjorkman & Zeijlstra 2019 and Deal 2022).

2.1 The inventory of polarity sensitive expressions

I begin by trying to establish a more complete inventory of polarity sensitive expressions in Irish than has so far been attempted. The element *tada*, illustrated in (5), is one such.

(5) a. Ní-or ith mé tada ar maidin.
C.NEG-PAST eat.PAST I anything/nothing on morning
'I didn't eat anything this morning.' 'I ate nothing this morning.'

[4]See Bauer 1997: p. 298–289, §19.5: 'Negative quantifiers do not exist in Maori. Sentence negation or other lexical means are used.' Mandarin, Korean and Bengali are also reported to lack such expressions (see Zeijlstra 2008: 15).

b. *D' ith mé tada ar maidin.
PAST eat.PAST I anything/nothing on morning

Tada (and its dialectal variants *dada* and *dadaidh*) is one member of a large class of elements united by a shared distribution and a shared interpretive profile. Like *tada*, they express existential quantification in the scope of sentential negation and are ill-formed when not so licensed. The illustrative examples below have the candidate elements in the scope of sentential negation, but readers should assume that the corresponding example without negation is ungrammatical. This is true for almost every case considered. Elements for which it is not (entirely) true will be discussed in Section 2.3 below. The larger class of environments in which PSEs appear (which do not all involve the explicit appearance of negation) will be considered in Section 2.2.

The class of PSEs of this type includes nominals, along with some adverbs of temporal perspective. The nominal group in turn includes two subtypes: a class of monomorphemic lexical items that stand alone (as in 5) and more complex nominal expressions that include one of a set of functional elements that define the containing expression as a PSE. One element of the latter type is the prepositional phrase *ar bith* (diachronically or literally 'in the world' or 'on earth'), which we have already encountered in (4a), and which is exemplified further in (6). This is a post-nominal modifier which attaches to indefinites in certain environments and converts its host into a PSE. Note that since negation is high in Irish (marked on C in finite clauses), elements in subject position will always be in the semantic scope and in the syntactic domain of negation when it is present (Acquaviva 1996; Duffield 1995; McCloskey 1996; McCloskey 2001; McCloskey 2017).[5]

(6) a. Ní-or thóg bean ar bith de na mná seo riamh
C.NEG-PAST take.PAST woman any of the women DEMON ever
an fiabhras.
the fever
'None of these women ever contracted fever.' GOG 132

[5]An anonymous reviewer raises the interesting question of whether or not the alternative word in Irish for the numeral 'one' (*amháin*) may also determine a PSE. The answer is that it does not. The two elements differ in a number of ways from one another: *aon* is pre-nominal, *amháin* is post-nominal. *Amháin* has a use as a focus particle meaning 'only', while *aon* does not. The question of why and how the two elements differ in their ability to determine a PSE clearly deserves a fuller investigation than is possible here and almost certainly has implications for the fundamental question of what makes a PSE. I am grateful to the reviewer for raising this important issue, even if these remarks in response must remain unsatisfying.

b. Níl taibhsí ar bith ann agus ní raibh ariamh.
is.not ghosts any in.it and C.NEG.FIN be.PAST (n)ever
'There are no ghosts and there never were.' COC 156

The numeral *aon* 'one', in addition, can be used to determine polarity sensitive nominals and in this use (and only in this use) it may compose with mass nouns (see 7b-c) and, in some dialects at least, with plural count nouns (see 7d).

(7) a. Ní-or dúradh aon chuid de seo riamh go hoscailte.
C.NEG-PAST say.PAST.IMPERS one part of this (n)ever openly
'No part of this was ever said openly.' PNG 187

b. Ní-or thugais aon ghrá ceart riamh dom.
C.NEG-PAST give.PAST.S2 one love proper (n)ever to.me
'You never gave me any proper love.' ANNF 59

c. Ní raibh aon eolas eile ar a mhalairt.
C.NEG.FIN be.PAST one knowledge other on its alternative
'Nobody knew any different.' LAN 26

d. Ní raibh aon bháid ar an Oileán ag na chéad daoine.
C.NEG.FIN be.PAST one boat.PL on the Island at the first people
'The first people on the Island hadn't any boats.' LAN 27

Éinne (in western and southern varieties) is a fusion of *aon* with the noun *duine* 'person':

(8) Ní-or labhair éinne ariamh liom.
C.NEG-PAST speak.PAST anyone (n)ever with.me
'Nobody ever spoke to me.'

Among the elements which seem to lack any internal structure are *tada*/*dada*, *faic*, and *a dhath*, all of which correspond to English 'anything' or 'nothing'.[6]

(9) a. Ní dúrt faic.
C.NEG.FIN say.PAST.S1 anything
'I didn't say anything.' AGFC 181

[6]Such elements, while themselves monomorphemic, can be "strengthened" by addition of *ar bith* or minimizers of the type to be considered in the subsection which follows:

(i) Ní-or ghá dóibh tada ar bith beo eile a dhéanamh aríst go deo.
C.NEG-PAST need to.them nothing any alive other do.NFIN again ever
'They never had to do a single solitary thing ever again.' AFAP 33

b. Ní léann siad tada.
C.NEG.FIN read.PRES they anything
'They don't read anything.' ACED 267

c. Ní ba léir dó a dhath ariamh.
C.NEG.FIN COP.PAST clear to.him anything (n)ever
'Nothing was ever clear to him.' GFH 102

In Munster dialects, the element *puinn* is also available. It can appear prenominally as in (10a), or alone, as in (10b), with the meaning 'at all' or 'much':

(10) a. Ní-or fhág puinn bád riamh an t-oileán seo.
C.NEG-PAST leave.PAST any boat ever the island DEMON
'No boat ever left this island.' CFC 129

b. Ní fhaca puinn ina dhiaidh sin é.
C.NEG.FIN see.PAST.S1 any after that him
'I didn't see him much at all after that.' LGL 421

The polarity sensitive adverbs are adverbs of temporal perspective like *(a)riamh* 'ever, still', *choíche*, *go brách*, *go deo* 'ever, forever'. The first (*ariamh*) is seen in (6a, 6b, 7a, 7b, 9c) and (10a). The others are exemplified in (11).

(11) a. Cuirimse de gheasa ort gan aon fhear a choíche
put.PRES.S1 of injunctions on.you NEG.NONFIN any man ever
a phósadh.
VCE marry.NFIN
'I put you under an injunction not to ever marry any man.' SRNF 49

b. Ní-or mhaith leis a bheith go deo ag caint air
C.NEG-PAST good with.him be.NFIN ever PROG talk on.him
féin.
REFL.LOG
'He didn't like to ever be talking about himself.' LGL 109

Within this group there is also an extensive catalog of expressions conventionally taken to denote the lowest possible point on some scale, so-called 'minimizers'. *Smid* 'breath' refers to the tiniest sound audible, *deor* 'drop' to the smallest imaginable quantity of a liquid, *ceo* 'mist' to the most insubstantial thing, while *pioc* 'a pick' refers to the smallest measure imaginable.

(12) a. Ach ní-or labhair sé smid leofa.
but C.NEG-PAST speak.PAST he breath with.them
'But he didn't breathe a word to them.' SIF 33

b. Ní-or chaoin sé deoir ariamh ina shaol.
C.NEG-PAST cry.PAST he drop ever in.his life
'He never shed a tear in his life.' GDDR 166

c. Ach ní-or fhéad sé ceo a dhéanamh.
but C.NEG-PAST can.PAST he mist do.NFIN
'But he couldn't do a thing.' CC 79

d. Níl pioc fírinne i n-a cheann.
is.not pick truth.GEN in his head
'He is incapable of telling the truth.' BTFS 124

An Irish particularity (apparently) is that many minimizers are conventionalized disjunctions with the disjunction necessarily interpreted within the scope of negation:

(13) a. Ní-or ghéill sí ionga ná orlach don chiúnas.
C.NEG-PAST yield.PAST she fingernail or inch to.the silence
'She didn't yield an inch to the silence.'

b. Ní thabharfadh duine ná deoraí freagra air.
C.NEG.FIN give.COND person or exile/stranger answer on.him
'Not a soul would answer him.'

c. Ní raibh tásc ná tuairisc ar an chainteoir.
C.NEG-PAST be.PAST sign or report on the speaker
'There wasn't a sign or a trace of the speaker.'

d. Ní-or chorruigheadar lámh ná cos ariamh ar a son.
C.NEG-PAST move.PASTP3 hand or foot ever on-their-behalf
'They didn't ever lift a finger on their behalf.'

e. Ní-or fhág sé bonn bán ná pingin rua ag a clann
C.NEG-PAST leave.PAST he coin white or penny red at her family
sise.
she
'But he didn't leave a penny to her family.' CTP 4

There are certain other minimizers which should be mentioned as well. The prepositional phrase *dá laghad* 'of the least' acting as a post-nominal modifier may create one such:

(14) Níl aird dá laghad aici ar an mbeirt.
is.not attention of.the least at.her on the two.people
'She doesn't pay the slightest attention to the two of them.'

And in a slightly more colorful turn of phrase, we have *faic na fríde*. *Faic* is the monomorphemic element already discussed in this section, meaning 'nothing' or 'anything'. When modified by the possessor *na fríde* 'of the mite' it signifies 'the slightest/tiniest thing':

(15) Níl faic na fríde le déanamh acu.
is.not to do.VN at.them
'They have nothing whatever to do.'

Finally, some minimizers are based on now opaque metaphors. In the Irish of Conamara, for example, the expression *mac an éin bheo* 'the son of the living bird' refers to the smallest possible set of people:

(16) Ní raibh mac an éin bheo le feiceáil an tráth
C.NEG.FIN be.PAST son the bird.GEN living.GEN ASP see.VN the time
sin de mhaidin.
DEMON of morning
'There wasn't a single solitary person to be seen at that time of the morning.'

There will be little or nothing in this inventory to surprise those who have closely studied the inventory of negative polarity systems in other languages.

2.2 Beyond negation: The licensing environments

We have so far considered just one environment in which the PSEs described in the previous section may appear: in the scope of sentential negation. But their distribution is, in fact, much broader.[7]

[7]Before extending the investigation to licensing contexts beyond that of negation, we should note that PSEs are also licensed in the scope of the emphatic or 'demonic' negation studied by Ó Siadhail (1989: 326–331) and especially by D'Antuono (2023). In this construction a phrase is fronted to a position immediately to the right of the emphatic negators *diabhal* 'devil' or *dheamhan* 'demon' and the clause out of which the phrase is extracted is headed by the WH-complementizer. PSEs may be fronted (as in i) or appear in a clause-internal position (as in ii):

All of them, for instance, also appear in the scope of semi-negative expressions such as 'rarely' or 'hardly' – which in Irish are predicates that select clausal complements (nonfinite or finite):

(17) Is rí-annamh anois éinne acu a theacht abhaile.
COP.PRES very-rare now anyone of.them come.NFIN home
'It's very rare now for any of them to come home.' CAA 265

(18) Ar éigean a bheas tada le déanamh agat.
hardly C be.FUT anything PTC do.NFIN at.you
'You'll have hardly anything to do.' AA 215

The PSEs we are concerned with appear in fact in a broad range of environments, all of them characteristic of those in which NPIs have been shown to appear in other languages. Documenting this pattern and its breadth is important work, but it makes for tedious reading. I have therefore gathered the relevant data in Appendix One. What is shown there is that PSEs in Irish, with their characteristic existential interpretations, are licensed in the following range of environments:

1. In polar questions
2. In WH-questions (when rhetorical or when demanding exhaustive, rather than partial, answers)
3. In conditional clauses (realis and irrealis)
4. In equative clauses
5. In comparative and superlative clauses
6. In phrases and clauses introduced by the degree particle *ró-* 'too'
7. In certain temporal clauses introduced by *sul* 'before' or *nuair* 'when'

(i) Dheamhan freagra ar bith a thug sí orm.
demon answer any C.WH give.PAST she on.me
'Not an answer did she give me.' AA 240

(ii) Diabhal duine a thug aon aird orm.
devil person C.WH give.PAST any attention on.me
'Not a person paid any attention to me.' DGD 79

8. In the complement of adversative (and some implicative) predicates

9. In the restrictive clauses of universal quantification structures

This is a list which is familiar from decades of research on NPIs. Appendix One documents this distribution and also considers some particularities of the Irish patterns, having to do especially with the licensing potential of adversative and negative implicative predicates.

Irish PSEs may in fact appear in an additional environment which, as far as I have been able to tell, has not so far been identified for NPIs in other languages – in clauses introduced by (the equivalent of) *to the extent that*. This is illustrated in (19):

(19) To the extent that anyone ever believed this …
sa mhéid is gu-r chuir mé aithne ar bith air
in.the extent as C-PAST put.PAST I acquaintance any on.him
'to the extent that I got to know him at all' LSC 130

The licensing potential of such a context presumably arises from the implication of doubt or of disbelief that it conveys concerning the content of the complement clause.

In sum: the distribution of PSEs in Irish and the distribution of NPIs in other well-studied languages are parallel in strikingly complete and exact ways. And although real progress has been made in recent years on the question of what this broad range of environments might have in common in terms of their semantics (for overviews, see Giannakidou 2011, Chierchia 2013, or Homer 2020), the task of identifying any plausible syntactic commonality, one that might provide the basis for an agreement or concord relation, seems very challenging.

2.3 Available readings

It is a well-known, if not well-understood, property of NPI systems that a subset of the NPIs of a language may appear outside the licensing environments just listed, but with quasi-universal (or perhaps generic) rather than existential force. These are the so-called "free choice" readings of certain NPIs (Bolinger 1972; Horn 1989: 400 ff; Horn 2000; Chierchia 2013: Chap. 6). Roughly half of the 115 languages in the Haspelmath (1997) sample allow this option for some of their negative polarity items. Irish can be added to that subgroup. The ill-formed (5b) above, for example, is well-formed if the main verb is in conditional mood:

(20) D' íosfadh sé tada.
PAST eat.COND he anything
'He'd eat anything.'

Among the PSEs, all but the minimizers allow such readings. The facilitating environments are the familiar ones – modal contexts (as in 20) or the presence of a restrictive modifier such as a relative clause modifying the PSE ('subtrigging' in the sense of LeGrand 1975). This possibility is exemplified for the nominal PSEs in (21) and in (22). (21) illustrates the modal environment (generic in 21b); those in (22) illustrate the 'subtrigging' effect.

(21) a. Ceann de na hoícheanta sin go dtarlódh faic
one of the nights DEMON C happen.COND anything
'one of those nights when anything could happen' LGL 250

b. maidir le daoine bochta, tá rud ar bith sách maith dóibh.
as-for people poor be.PRES thing any enough good for.them
'As for poor people – anything is good enough for them.' CG 59

c. Dhéanfainn rud ar bith ach tusa a fháil domh féin.
do.COND.S1 thing any but you VCE get.VN to.me REFL.LOG
'I'd do anything to get you for myself.' SSOTC 222

(22) a. Rud ar bith a tugadh ar iasacht domsa ariamh thug
thing any C give.PAST.IMPERS on loan to.me ever give.PAST
mé ar ais é.
I back it
'Anything that I was ever lent, I gave it back.' AA 90

b. Aon áit a chuais, ní raibh aon ní á
one/any place C go.PAST.S2 C.NEG-PAST be.PAST any thing PROG.PASS
labhairt ach Gaolainn.
speak but Irish
'Any place you went, there wasn't anything being spoken but Irish.'
TMGB 39

The temporal adverbials *(a)riamh*, *go brách*, *go deo* and *choíche* 'ever' may, in addition, appear outside the licensing environments just discussed and in that context they have universal rather than existential force and translate as English 'always' or 'forever':

(23) a. Bhí sé ariamh ann.
be.PAST it ever in.it
'It has always existed.' FF 167

b. An síleann tú go bhfuil tú ag imeacht go brách uainn?
C.Q think.PRES you C be.PRES you PROG leave.VN forever from.us
'Do you think you are leaving us for ever?' ATFS 207

c. Bhíodh clocha i nGleann Easa riamh agus beidh go deo.
be.PAST.HABIT stones in ever and be.FUT always
'There have always been stones in Gleann Easa and there always will be.' DEAD 96

d. Beidh cuimhne choíche agam air.
be.FUT memory ever at.me on.it
'I will always remember it.' OMGS 290

The two meanings expressed by these adverbs are at least close to those expressed by NPIs and their "free choice" counterparts. In addition, both interpretations (existential in NPI-licensing contexts, universal otherwise) are available across the class, suggesting that something more systematic than lexical polysemy is at work. It may be, then, that the possibilities seen in (23) reflect a "free choice" option for certain NPIs. These universal readings, however, are not subject to the licensing restrictions observed for the determiner NPIs such as English *any* (LeGrand 1975; Kadmon & Landman 1993; Dayal 1998; Dayal 2013; Giannakidou 2001; Chierchia 2013: Chap. 6), for which modality or genericity seems to be crucial, neither of which is relevant for (23).

It might, alternatively, be more profitable to think about cases such as (23) as being parallel, in relevant respects, to cases such as English *until* and related items in other European languages. Temporal clauses introduced by *until* are strong NPIs when interpreted as punctual:

(24) a. He didn't finish the paper until July.
b. * He finished the paper until July.

But when modifying atelic predications they are, as seen in (25), durative in their interpretation and are positive polarity items expressing extension over relatively long intervals (Karttunen 1974; Mittwoch 1977; Giannakidou 1992; Declerck 1995; de Swart 1996; and especially Condoravdi 2008):

(25) We remained in Cambridge until the end of the year.

The connection with our Irish cases is that the non-NPI readings of the temporal adverbs in (23) also appear only in the context of atelic predications.[8] Needless to say, this discussion is little more than a marker laid down for a future research project.[9]

2.4 Non-local licensing

One of the principal themes of research on negative concord systems has been that of locality. To a first approximation, the concord relation may not cross a finite clause boundary unless it is subjunctive (Haegeman & Zanuttini 1991, Zeijlstra 2008: 43–45, Deal 2022).[10] This is why analyses of the concord relation in terms of agreement or movement, with their associated locality requirements, have been persuasive and influential. The licensing of PSEs in Irish, however, is subject to no such restriction, as shown by examples like those in (26–28), which are commonplace and frequent (304 in our data-set).

(26) a. Ní-or chualas gu-r mharaigh na tramanna duine
C.NEG-PAST hear.PAST.S1 C.PAST kill.PAST the trams person
ar bith ariamh.
any ever
'I haven't heard that the trams ever killed anyone.' CTP 49

b. má cheapann sibh go bhfuil mise ag déanamh aon fhocal
if think.PRES you.PL C be.PRES I PROG make.VN any word
bréige
lie.GEN
'if you think that I am telling any lies' MABAT 56

c. An síleann tú go dtiocfadh le cailín ar bith grá mar sin a
C.Q think.PRES you C could girl any love like that VCE
thabhairt uaithe?
give.VN from.her
'Do you think that any girl could give such love?' ATFS 343

In each of the cases in (26), the licensing element (negation, the complementizer *má* in (26b) or the polar interrogative particle in (26c)) is separated from the PSE it

[8] The same restriction seems to hold for the more or less archaic use of English *ever* when universal in its force (see Israel 1998; Horn 2000: 181–183).

[9] I am grateful to Nicola D'Antuono for discussion of these matters.

[10] For complications, exceptions, and for approaches to those issues, see Robinson & Thoms (2021), among others.

licenses by at least one finite CP-boundary. Longer dependencies are also possible, as in (27). In (27a), the licensed PSE is separated from its licensing negation by two finite clause boundaries, one of them the complement to the experiencer noun *súil* 'hope'; in (27b) the licensing environment is established by the noun *eagla* 'fear', which is adversative and licenses PSEs in its complement.

(27) a. Ní hamhlaidh [$_{CP}$ a tá súil agam [$_{CP}$ go dtiocfaidh
COP.NEG so C be.PRES hope at.me C come.FUT
éinne]].
anyone
'It is not the case that I really expect that anyone will come.' LGL 60

b. ar eagla [$_{CP}$ go gceapfadh sé [$_{CP}$ go raibh duine ar bith díobh
on fear C think.COND he C be.PAST person any of.them
chomh díthcéillí]]
so foolish
'for fear that he would think that any of them were so foolish' ATIM 90

The licensing relation can also span at least some island-boundaries, as seen in (28), a WH-island in (28a)[11], and in (28b) a finite CP-argument of the noun *cuma* 'appearance'. Such structures are strong islands in Irish (McCloskey 1985; McCloskey 2002; Maki & Ó Baoill 2011.)

(28) a. Cha-r fhoghlaim mé ariamh cén dóigh le rud ar bith a
C.NEG-PAST learn.PAST I ever what way with thing any VCE
tharraingt.
draw.VN
'I didn't ever learn how to draw anything.' APB 12

b. Ní raibh cuma uirthi go raibh eagla ar bith roimh
C.NEG.FIN be.PAST appearance on.her C be.PAST fear any before
an astar uirthi.
the journey on.her
'It didn't look as if she had any fear of the journey.' NLAB 54

[11]For (28a) one might wonder whether the PSE is licensed in the interrogative clause itself rather than by the matrix negation. This is not a plausible interpretation, though, given that the WH-interrogative is neither rhetorical nor exhaustive. In addition, the example is ill-formed when the matrix negation is removed.

2.5 Modification by 'almost'

Finally, all PSEs in Irish strongly resist modification by *almost*, a property which in many languages distinguishes NPIs from inherently negative expressions:

(29) a. *Ní raibh comhair a bheith duine ar bith i láthair.
C.NEG-PAST be.PAST almost person any present
'There was almost nobody present.'

b. *Ní dhéanainn freastal ar chomhair a bheith
C.NEG-PAST do.PAST.HABIT.S1 attendance on almost
léacht ar bith.
lecture any
'I attended almost no lectures.'

2.6 Interim conclusion

The items surveyed in this section (all of the PSEs so far identified in Irish), while being polarity sensitive: (i) are incapable of expressing negation outside an appropriate licensing context, (ii) appear in the range of environments typical of NPIs investigated in other languages, (iii) support quasi-universal readings outside those environments, (iv) can be licensed non-locally, even across certain island boundaries, and (v) are incompatible with modification by *almost*. All of this suggests that Acquaviva (1996) was correct in arguing that Irish possesses a rich and familiar inventory of negative polarity items but has no plausible candidate for the role of inherently negative expression, as far as is known at present. It therefore also lacks the mechanisms of negative concord and our PSEs are negative polarity items.[12] From this point on, then, I will abandon the neutral term PSE and call all of the items discussed here negative polarity items (NPIs).

In light of that conclusion, however, the observations of the section which follows seem unpleasantly anomalous.

[12] Elena Herburger (2001) develops an important analysis of the distribution of polarity sensitive expressions in Spanish, another case in which the distinction between NPIs and INEs seems less than clear. She shows that that complex of data can be accounted for on the assumption that the relevant PSEs in Spanish are systematically ambiguous between being NPIs and items lexically specified as being 'negative'. Her analysis is remarkably successful, but it cannot be applied to the problems we deal with in the following section. In Spanish, the class of elements Herburger examines can always express negation on their own, so to speak (because they are inherently negative expressions). But that is not possible for the class of Irish elements we are concerned with here, as we have seen with examples like (5b) above. Put differently, the set of contexts in which the '*n*-words' of Spanish may appear is the union of the distributions of NPIs and what I have called here INEs, a distribution much broader than that of the Irish PSEs we are concerned with.

3 The anomaly

Each of the polarity sensitive elements identified in the previous section may appear in apparent isolation as a subsentential fragment, often in answer to a WH-question (as in 30 or 2 above) or to a polar question (as in 31).[13,14]

(30) NPIs AS FRAGMENT ANSWERS TO WH-QUESTIONS:

a. "Céard a tá uait?" "Tada, a Mháistir."
what C.WH be.PRES from.you anything, VOC-PTC Master
'"What do you want?" "Nothing, sir."' LL 254

b. "Agus caidé a ghní tú leis na réaltógaí?" "Rud ar bith."
and what C.WH do you with the stars thing any
'"And what do you do with the stars?" "Nothing."' APB 47

c. "Cén chúis a gcuirfeá an cheist sin orm?" "Ó, cúis ar bith."
what reason C put.COND.S2 the question DEMON on.me oh reason any
'"Why would you ask me that question?" "Oh, no reason."' LOFRS 241

d. ach cé labhrann liom í? Éinne ach do leithéidse.
but who speak.PRES with.me it anyone but your like
'But who speaks it to me? Nobody except the likes of you.' TMGB 252

(31) NPIs AS FRAGMENT ANSWERS TO POLAR QUESTIONS:

a. "An ndéanfaidh aon duine m' áit-sa duit?" "Go deo."
C.Q make.FUT any person my place for.you ever
'"Will anyone ever take my place for you?" "Never."' ATFS 488

b. "An bhfaigheadh na scéalaithe aon díolaíocht?" "Aon rud in aon chor."
C.Q get.PAST.HABIT the storytellers any payment any thing at.all
'"Would the storytellers get any payment?" "Nothing at all."' AL 88

c. "Agus ní raibh eagla ort roimhe?" "Eagla ar bith."
and C.NEG.FIN be.PAST fear on.you before.him fear any
'"And you weren't afraid of him?" "Not at all."' D 18

[13] Gary Thoms reports that similar facts hold for Scottish Gaelic.

[14] Example (30b) is from a translation of *Le Petit Prince* by Antoine de Sainte-Exupéry. The French original has: *Et que fais-tu de ces étoiles? Rien. Je les possède.*

It is perhaps worth emphasizing that the answers in (30) and (31) are in no way strained. They require no particular contextualization, nor do they demand any special accommodation. They are routine, and as far as I am aware, there is no alternative way to express what they express (fully articulated clauses aside).

There are additional contexts in which such NPI fragments appear, contexts that do not (in an obvious way at least) involve question-answer pairings. We will return to those cases and their implications, but the dilemma is already clear. If the arguments developed so far in this chapter are to be relied upon, bare NPIs may function as fragment answers in Irish. But there is very strong evidence from a range of languages already studied that NPIs cannot serve as fragment answers. In fact, this has come to be recognized as one of the most reliable diagnostics for distinguishing between NPIs and inherently negative expressions. As Penka & Zeijlstra (2010: 778) put it:

> The ability to contribute negation in fragmentary answers can, thus, be regarded as a defining property of negative indefinites, distinguishing them from NPIs (cf. Bernini and Ramat 1996 and Haspelmath 1997).

We might, in the face of this dilemma, reject the arguments of the first half of this chapter and conclude that all Irish polarity sensitive expressions are actually negative indefinites. My own assessment (unsurprisingly) is that this would not be a wise move and in the remainder of the chapter, I propose an analysis which preserves those earlier results, while also preserving the basic integrity of the generalization articulated by Penka and Zeijlstra. The resolution proposed is well-supported by evidence internal to Irish, and it also has interesting theoretical implications. The questions which need to be addressed in such a resolution are these:

- How can the fragments in (30–31) be well-formed outside the licensing context that they otherwise require?
- How can such examples have the interpretations that they do in the absence of that crucial licensing environment?
- Why are the mechanisms, whatever they may be, that underlie the possibility in (30–31) unavailable in the other languages so far examined?

The answer to all of these questions, I believe, is that Irish has a movement rule whose very particular properties are crucial in permitting the possibilities on display in (30–31).

4 The resolution

4.1 Narrative fronting

Ó Siadhail (1989: §9.2.2) and McCloskey (1996) discuss a process, named by Ó Siadhail *Narrative Fronting*, by way of which a phrase is moved leftward in a finite clause to a position immediately to the left of the negative complementizer and, therefore, also to the left of the inflected verb. The process is productive and of high frequency.[15]

A variety of phrase-types may be fronted, but the most frequently attested pattern is one in which the moved phrase is an indefinite nominal within the scope of the triggering negation. The fronted phrase is nominal in (32), adverbial or prepositional in (33).

(32) NARRATIVE FRONTING IN NEGATIVE CLAUSES:

a. Duine níba réasúnaí ní raibh ann.
person more reasonable.COMPR NEG was in.it
'A more reasonable person was there none.' FF 107

b. Mo bhéal ní-or oscail mé ar feadh chúig lá.
my mouth C.NEG-PAST open.PAST I during five day
'I didn't as much as open my mouth for five days.' TUAIR 26-04-21

c. Leabhar gramadaí ní raibh ariamh agam.
book grammar.GEN C.NEG-PAST be.PAST ever at.me
'I never had a grammar-book.' ABHM 41

(33) a. Ach díreach ní-or bhreathnaigh sí air
but straight C.NEG-PAST look.PAST she on.him
'But straight she didn't look at him.' C 24

[15]There is a similar but much less productive process that applies in clauses headed by the complementizer *go* and in which the inflected verb is in subjunctive mood (a form now archaic for almost all speakers). Such clauses express curses (as in ia) or blessings (as in ib):

(i) a. Na seacht ndiabhal déag go dtuga _ leo sibh!
the seven devil ten C take.SUBJ with.them you.PL
'May the seventeen devils take you!' CNF 51

b. Ádh agus sonas go raibh _ ort.
luck and happiness C be.SUBJ on.you
'May you have good fortune and happiness.' SK 104

Though superficially similar, it is not clear that the two processes have a common syntax. We will be concerned here only with the very productive negative case.

b. isteach san fháinne ní thiocfaidh sí
into in.the ring C.NEG.FIN come.FUT she
'Into the ring she will not come.' SGC 112

c. Go deo ná go bráthach ní scarfamaoid ón a chéile arís.
ever or ever C.NEG.FIN separate.FUT.P1 from each-other again
'Never again will we separate from one another.' IAE 331

Narrative Fronting is optional and the examples of (32) and (33) are equivalent in their truth-conditions to the corresponding examples in which it has not applied. (32) and (33), though, are felt to be "emphatic" in a way that their counterparts without fronting are not. The general pattern, then, can be schematized as in (34):

(34) NARRATIVE FRONTING:

a. [XP_j C V ... $\__j$ ...]
[NEG] [FIN]

b. where XP can be any constituent type but is frequently an indefinite nominal

c. and the interpretive effect is to express "emphatic" negation.

Narrative Fronting is relevant for us because all of the weak NPIs and all of the minimizers surveyed earlier appear freely and frequently in the XP-position of (34a). The examples in (35) illustrate this fact for the weak NPIs; those in (36) for the minimizers.

(35) WEAK NPIS IN NARRATIVE FRONTING:

a. Cearta ar bith ní raibh _ ag gnáthdhaoine.
rights any C.NEG.FIN be.PAST at ordinary.people
'Ordinary people had no rights. ABHM 53

b. Aon mhoill ní-or dhein Cromaill. _
any delay C.NEG-PAST make.PAST Cromwell
'Cromwell made no delay.'/ 'No delay did Cromwell make.' OOGC 199

c. Ach tada ní raibh sé in ann _ a chloisteáil.
but anything C.NEG.FIN be.PAST he able VCE hear.VN
'But nothing was he able to hear.' SJSJ 55

d. Go deo arís ní dhéanfadh fear amadán _ dom
ever again C.NEG.FIN make.COND man fool of.me
'Never again would a man make a fool of me.' LGL 158

e. Éinne de na comharsain ní cheannódh é.
anyone of the neighbors C.NEG.FIN buy.COND it
'None of the neighbors would buy it.' BM 197

(36) MINIMIZERS IN NARRATIVE FRONTING:

a. Faic na fríde ní bhfuair mé _ mar fhreagra.
the-tiniest-thing C.NEG.FIN get.PAST I as answer
'I didn't get the tiniest thing as an answer.' PAA 24

b. Smid ní -l _ ann faoi Tone.
breath C.NEG.FIN be.PRES in.it about
'There's absolutely nothing in it about Tone.' TII 121

c. Pioc eagla ní raibh _ ar an tiománaí.
pick fear C.NEG.FIN be.PAST on the driver
'The driver wasn't the tiniest bit afraid.' DR 15

d. Deoir ní-or chaoin sé. _
drop C.NEG-PAST weep.PAST he
'He didn't cry a single tear.' LL 118

e. Le mac an éin bheo níor sceith ceachtar
with son the bird.GEN living C.NEG-PAST expose.PAST either
againn ár rún.
of.us our secret
'Neither of us revealed our secret to a single living soul.' LL 437

It is not just that weak NPIs *may* undergo Narrative Fronting; they clearly have a particular affinity for the environment created in (34), as is shown by the fact that in 43% of the attested examples of Narrative Fronting in our database (196 of 486), the element fronted is an NPI. Among these, minimizers are particularly frequent; they represent 41% (80 of 196) of all the examples in which NPIs are fronted under Narrative Fronting.

There is something, then, about the environment of Narrative Fronting that particularly favors NPIs and there is something about NPIs (and minimizers, in particular) which makes them especially susceptible to fronting in this context. This is one of a number of observations suggesting that the two phenomena are deeply entangled. What is the nature of that entanglement though?

The "emphatic" character of Narrative Fronting seems to have its source, at least in part, in that the structure in (34) expresses "scalar assertions" in the sense of Krifka (1995). They evoke scalar implicatures of a familiar kind in that the use of such a structure evokes alternatives to the proposition actually expressed – alternatives that are ranked on a scale of strength. The relevant notion of "strength" is in turn dependent on information shared among interlocutors, but also on the relative informational strength of those alternatives as measured by asymmetric entailment relations. Such ranked alternatives, implicitly evoked, have been central to theoretical work in pragmatics and semantics for many years. In that light, consider the examples of (37):

(37) a. Acht sagart amháin ní tháinig de chóir fhéasta an Rí.
but priest one C.NEG.FIN come.PAST near feast the king
'But not a single priest came near the king's feast.' UMI 13

b. Míle fear ní bhainfeadh feacadh aisti as a háit.
thousand men C.NEG.FIN take.COND movement out.of.it from its place
'A thousand men couldn't budge it from its place.' (a rock) ATT 34

In (37a) the alternatives evoked have to do with the number of priests who had attended the king's feast – a set of propositions of the form: '*n* priests did not attend the king's feast' ordered by the value of *n*. The asserted proposition is that the lowest number possible (namely none at all) attended. That is also the strongest proposition among the evoked alternatives because it entails all of the others (i.e., if it is not the case that one priest attended, then it is not the case that two attended, or that three attended, or four or …). The proposition expressed is therefore the logically strongest and the most informationally specific of the alternative-set. In (37b), the alternative propositions evoked are of the form: '*n* men could not move that rock', and those propositions are ranked by the value of *n* from one thousand down as far as one. The strongest proposition on that scale (in the sense of entailing all of the others and of being, again, informationally more specific) is the one asserted to be true (i.e., if it is impossible for one thousand men to dislodge the rock, it is impossible for 900 to do so and also 800 and so on downwards). The meaning ultimately conveyed, then, is that the rock is likely impossible to dislodge. The proposition actually expressed in both cases is presented as being at the extreme high-point of a scale of salient alternatives. This is the standard logic of scalar implicatures. It seems reasonable, then, to assume

that Narrative Fronting is a syntactic operation which attracts to the XP-position of (34) constituents that evoke alternatives. (38) revises (34) accordingly:

(38) NARRATIVE FRONTING:

a. $[\ \mathrm{XP}_j \quad \underset{[\textsc{neg}]}{\mathrm{C}} \quad \underset{[\textsc{fin}]}{\mathrm{V}} \quad \ldots \ -_j \ldots\ ]$

b. where XP is alternative-evoking.

If the phrase fronted in (38) must be alternative-evoking, an additional property of Narrative Fronting falls into place. That phrase is modified very frequently by the focus particle *féin*, a phrase-final focus marker whose meaning is close to that of English 'even':

(39) Bhí an madra féin fachtha mífhoighneach.
be.PAST the dog *féin* gotten impatient
'Even the dog had become impatient.' GSA 19

The particle *féin* appears very frequently as a modifier of the phrase fronted under Narrative Fronting:

(40) a. An toirneach féin ní dhúiseodh Johnny.
the thunder *féin* C.NEG.FIN waken.COND
'Even thunder wouldn't waken Johnny.' C 17

b. An fhuiseog féin ní raibh ina suí.
the lark *féin* C.NEG.FIN be.PAST awake
'Even the lark wasn't awake.' AN 43

c. Feoirling féin ní thabharfadh sé dhó.
farthing *féin* C.NEG.FIN give.COND he to.him
'Even a farthing he wouldn't give him.' ATIM 123

This is an expected possibility given (38) because the effect of suffixing *féin* to some phrase XP is exactly to turn XP into an alternative-evoking expression. In the case of (40b), for instance, the alternative propositions evoked have to do with what creatures were up and about that morning. Given conventional ideas about bird life-styles, the lark will always be the earliest creature awake and the proposition that the lark was not awake therefore entails all of the alternative propositions evoked (i.e., the fox was not awake, the hare was not awake, the curlew was not awake …). What is ultimately conveyed, then, is that no creature stirred and that was because it was really unusually early in the morning.

Such assertions, then, convey that the proposition actually expressed is at the upper limit of some scale of imaginable alternatives and is therefore outside the range of conventional norms and expectations, thereby evoking in hearers a sense of surprise or unexpectedness. This seems to be the principal source of the intuitively "emphatic" character of Narrative Fronting examples; they are scalar assertions in Krifka's (1995) sense.

Negation plays a central role in these deductions. Its effect is to reverse the direction of entailment among the alternatives and therefore to reverse the ranking of those propositions on the scale of strength. Returning to example (37a), for instance: the proposition *One priest attended the king's feast* leaves open the possibility that two priests, or three, or four … attended the feast, but it entails neither those propositions nor their negations. The presence of negation changes that calculation and rules out possibilities that would be allowed in its absence; it therefore 'converts' what would be in its absence a logically weak and low-ranked proposition into a logically strong and high-ranked proposition.

This must be why the NPIs which are our central concern are so susceptible to Narrative Fronting.

If there were a class of expressions lexically specified to be alternative-evoking, we would now expect those expressions to appear naturally and frequently in the XP-position of (38). But that is exactly the claim that is at the heart of one of the most important strands of current research on the nature and licensing of NPIs – from the domain-widening of Kadmon & Landman (1993) to the explicit appeal to scalar implicatures in the work of Krifka (1995), Lee & Horn (1994), Israel (1998), Lahiri (1998), Horn (2000), Condoravdi (2008), Chierchia (2013) and Jeong & Roelefsen (2023). The central commitment in this line of work is that the limited distribution of NPIs (that is, that they may appear only in downward-entailing environments) is to be attributed to the fact that they are required in their lexical semantics to be alternative-evoking and further that they are, or at least that many of them are, lexically specified as representing minimal elements on the quantity-scale implicitly defined by those alternatives. But by the logic we just reviewed those "minimal" elements will be strengthened exactly when they appear within the scope of negation. Different theoreticians have different ways of working this reasoning into a formal theory of NPI-licensing, but the common thread, since Krifka (1995), has been that unless such items appear within the scope of negation (or an element with similar logical properties) they run afoul of a version of Grice's (1975) Quantity Maxim (namely, "be as informative as possible") built into the compositional mechanisms.

But negation is, of course, also the syntactic driver of Narrative Fronting. Weak NPIs and minimizers will be ideal candidates for the role of XP in (38), then, be-

cause they are in their lexical definition alternative-evoking and minimal. In their interaction with negation, then, they will very naturally generate the logically strong propositions that are the hallmark of the construction.

Given this perspective, we understand why NPIs in Irish appear so frequently in the fronted position of Narrative Fronting structures. The two phenomena are "entangled", as I put it above, because they exploit the same logical mechanism in evoking scalar implicatures – namely, the strengthening effect of downward entailing contexts. The difference between the two is that Narrative Fronting is syntacticized and so limited to a single downward-entailing environment (the domain of sentential negation), while weak NPIs and minimizers can exploit that logic in any environment that has the appropriate semantics (those described in Section 2.2). The descriptions given earlier now reduce to (41).

(41) NARRATIVE FRONTING:
The finite negation head may include a probe that attracts XPs that are alternative-evoking.

The description in (41), unlike our earlier formulations, makes no mention of the "emphatic" character of Narrative Fronting. That aspect of the construction, as we have seen, emerges organically from the interaction between the semantics of the negative head and the alternatives evoked by the attracted constituent.

A property of this account that is neither obviously correct nor obviously incorrect is that the fronting itself plays no role in establishing the "emphatic" character of Narrative Fronting. On this account, that aspect of the construction emerges from a semantic-pragmatic interaction between the attracting negation and the alternatives evoked by the fronted element; an interaction that would take place even when the alternative-evoking phrase remains in its base position. The pairs of examples in (42) and (43) should then be equally "emphatic":

(42) a. Ach ní-or labhair sé smid leofa.
but C.NEG-PAST speak.PAST he breath with.them
'But he didn't breathe a word to them.' SIF 33

b. Smid níor labhair sé _ leofa.

(43) a. Aon mhoill ní-or dhein Cromaill. _
any delay C.NEG-PAST make.PAST Cromwell
'No delay did Cromwell make.' OOGC 199

b. Níor dhein Cromaill aon mhoill.

Assessing whether or not this prediction is correct is a matter of such subtlety and vagueness that it will be next to impossible to investigate responsibly, I suspect. For what it may be worth (not much), my impression is that it is not obviously incorrect.[16]

4.2 Narrative fronting and scope

There is a final property of Narrative Fronting that we need to attend to, one that is implicit in the preceding discussion but which should be made explicit.

It was observed in McCloskey (1996: 76–86) that Narrative Fronting does not expand the scope of the raised item, at least with respect to negation. In (44), for instance, the temporal indefinite *uair amháin* remains within the scope of negation despite apparently preceding it:

(44) uair amháin fiú, ní-or cheistigh sé conas a bhí an
time one even C.NEG-PAST question.PAST he how C be.PAST the
t-ullmhúchán faoi bhráid Mheiriceá ag dul.
preparation for America PROG go.VN
'Not even once did he ask how the preparation for America was going.'
PI 67

(37a) above is similar, as are the examples in (45):

(45) a. Aon mhála amháin ní bhfaighidh tú.
one bag one C.NEG.FIN get.FUT you
'Not one bag will you get.' OTA 194

[16]Note that to claim that certain expressions are alternative-evoking is not to claim that they are "F-marked" in the sense familiar from work on the distribution of focal accents and contrastive focus (or to claim that Narrative Fronting is movement to a dedicated focal position). Rather I follow Krifka (1995), Jeong & Roelefsen (2023) and others in assuming that expressions in focus evoke alternatives but that that is just one of the contexts in which alternatives play a central role. See Krifka (1995) and Jeong & Roelefsen (2023) for important discussions of the issues that arise here. If we were to assume that the alternative-evoking character of NPIs reflects a kind of inherent focus marking, we are left in a poor position to understand the differences between emphatic and non-emphatic uses of NPIs – the central concern of the discussion in Jeong & Roelefsen (2023). They assume, with Krifka (1995) and others, that all NPIs evoke alternatives, as a matter of lexical specification, but that there is also a distinction between contingently-emphatic and inherently emphatic members of the class. The latter are the minimizers and they are inherently focused; the former may or may not be focused. These commitments are entirely consistent with the discussion here. Chierchia (2013) draws the same distinction in a different way. The minimizers, because they rely on an operator like *even* for the required exhaustification of their alternative-set, are always and strongly emphatic. For run of the mill weak NPIs, like *any*, however, the activation of the relevant alternatives is often undetectable.

b. duine amháin as ocht nduine dhéag i seomra na nuachtóirí
person one out.of eight person ten in room the journalists
ní-or labhair liom air
C.NEG-PAST speak.PAST with.me on.it
'Not one person out of eighteen in the newsroom spoke to me about it.' AAG 027

c. Duine acu ní-or aithníos.
person of.them C.NEG-PAST recognize.PAST.S1
'I recognized none of them.' BTFS 14

(45c), for example, in its actual context of use, does not convey that there was one person that I did not recognize – a meaning naturally expressible in English by means of *One of them, I didn't recognize*. Rather, the intended interpretation has the indefinite interpreted within the scope of negation. The examples of (46) show that fronted disjuncts also remain within the scope of the negation which triggers Narrative Fronting.

(46) a. Do chlann ná do chéile ní fheicfidh tú go deo.
your family or your spouse C.NEG.FIN see.FUT you ever
'You will never see either your spouse or your family.' CDC 230

b. Agus mionnán ná bainne ní bheadh … an bhliain sin
and kid.goat or milk C.NEG.FIN be.COND the year DEMON
aici.
at.her
'And she would have neither a goat nor (its) milk that year.' CC 16

For detailed discussion see McCloskey (1996).[17]

The fact, then, that weak NPIs and minimizers may appear in clause-initial position under Narrative Fronting (as in 35 and 36) is just one aspect of this larger generalization. In cases like (35) and (36), the fronted NPI remains within the scope of sentential negation just like the indefinites of (45) and the disjuncts of (46), and this is why such examples are fully well-formed.

[17]Note that the claim is not that wide-scope indefinites are never found in the fronted position of a Narrative Fronting structure. They are, though rarely. That is, they are found with the same degree of difficulty and at the same (low) level of frequency as is characteristic of indefinites in the base-position of the movement. That is, Narrative Fronting does not expand or change the scopal properties of fronted elements with respect to negation.

4.3 Narrative fronting and ellipsis

If the conclusions of the previous subsection are safe, the path is clear towards resolving the apparent anomaly I opened with: how there can be answers like (2), which consist only of a negative polarity item, in apparent violation of an otherwise valid crosslinguistic generalization. Following Merchant (2004: 691), we can maintain that the possibility of fragment answers consisting of, or containing, NPIs is parasitic on a prior application of Narrative Fronting.[18] A routine application of Narrative Fronting raises the NPI to the left periphery, and the clausal remnant from which it has been raised is, if conditions warrant, elided by a sluicing-like operation that eliminates all but the fronted phrase. The apparently isolated NPIs are well-formed because they are within the scope of negation in a pre-ellipsis representation, and they have the interpretations that they do because of the scope-preserving property of Narrative Fronting. Many languages disallow the equivalents of (30–31) because they lack a movement operation with the particular set of properties that we have demonstrated for Narrative Fronting in Irish (though see Laka 1993 and Giannakidou 2000 for cases of the same general type). In (47b), which involves a minimizer, L's response will be derived roughly as in (48b), in which I use a greyed-out font to indicate elided material.

(47) a. J: Ná habair a dhath le n-ár gcairde.
C.NEG.FIN say.IMPV anything with our friends
'Don't say anything to our friends.'

b. L: Smid.
breath
'Not a word.'

(48) a. Smid ní dhéarfaidh mé le n-ár gcairde.
breath C.NEG.FIN say.FUT I with our friends
'I won't breathe a word to our friends.'

b. i. ní dhéarfaidh mé smid le n-ár gcairde.
ii. smid ní dhéarfaidh mé _ le n-ár gcairde.
iii. smid [ní dhéarfaidh mé _ le n-ár gcairde]

[18]The link between Narrative Fronting and the possibility of fragment NPIs is made independently in D'Antuono (2024).

There is an additional set of observations that shows that the link between Narrative Fronting and legal fragment-types is very tight. There are items in Irish which resemble NPIs in many respects but to which Narrative Fronting may not apply. One of these is a focus-exceptive construction, which is one of the ways of expressing *only*. In these structures, which have much in common with the *ne-que* construction of French, the exceptive particle *ach* attaches to a focused phrase (or to a larger phrase which includes a focused phrase) within the domain of sentential negation:

(49) a. Ní-or labhair ach seisear.
C.NEG-PAST speak.PAST but six.people
'Only six people spoke.'

b. Ní ólaim ach tae.
C.NEG.FIN drink.PRES.S1 but tea
'I drink only tea.'

c. Ní raibh gluaisteán ach aige daoine saibhre an uair sin
C.NEG.FIN be.PAST car but at people rich the time DEMON
'Only rich people had cars at that time.' ABFS 26

Ach-phrases like those in (49), interpreted like 'only', are licensed in some, but not quite all, of the contexts in which NPIs are licensed. They appear, for instance, in the scope of negation or in polar questions:

(50) a. nó an raibh ann acht rud a samhladh dó
or C.Q was in.it but thing C imagine.PAST.IMPERS to.him
'Or was it only something he imagined?' LCS 110

b. An raibh ach an t-aon Naomh amháin ina measc go léir?
C.Q was but the one saint one in-their-midst all
'Or was there only one Saint among them all?' AG 113

Despite their kinship with NPIs (discussed in McCloskey 2013) such exceptives are excluded from Narrative Fronting structures:

(51) *Ach do dheartháir ní raibh ag an chruinniú aréir.
but your brother C.NEG.FIN be.PAST at the meeting last.night
'Only your brother was at the meeting last night.'

They are also correspondingly impossible as fragments, as shown in (52):

(52) a. Cé a bhí ag an chruinniú aréir?
who C be.PAST at the meeting last.night
'Who was at the meeting last night?'
b. *Ach do dheartháir.
but your brother
'Only your brother.'

Such close correlations are expected on the account developed here.[19]

4.4 English Redux

Before we move on to larger questions, there is a final empirical issue that should be dealt with: this chapter opened by announcing the goal of better understanding a claimed contrast between English and Irish – that Irish does, but English does not, allow fragment NPIs. There is, though, a strand of research that questions the claim for English – or at least suggests that the facts are more nuanced (den Dikken et al. 2000; Valmala 2007; Weir 2014; Weir 2015). At issue is what we should conclude about the example in (53):

(53) Q: What DIDN'T Owen buy? A: Any wine.

As all investigators have been careful to note, (53) is not accepted by all speakers and is, for many or most, of intermediate acceptability (I know of no quantitative study). The conditions that allow the fragment NPI in (53) seem also to be quite stringent; as noted by den Dikken et al. (2000: fn3, 44–45) and Weir (2014: 167–171), what is required is a negative WH-question with verum focus and a discourse context that includes, implicitly or explicitly, a set of propositions like that in (54):

(54) Owen bought pizza.
Owen bought bottled water.
Owen bought beer.
Owen bought chips.

The possibility in (53) also seems to be linked to the possibility of specificational pseudo-clefts in English like those in (55), in which the appearance of the NPI within the pivot position is also puzzling on most accounts.

(55) a. What we didn't make was (we didn't make) any progress.
b. What Owen didn't buy was (he didn't buy) any wine.

[19]Speakers often use Narrative Fronting in providing paraphrases for fragment answers.

It is exactly this connection that den Dikken et al. (2000) are centrally concerned with and in pursuing that connection they are brought to the conclusion that (53) is an elided form of (56):

(56) Q: What DIDN'T Owen buy? A: [He didn't buy] any wine.

The connection with the pseudoclefts in (55) is then that the relation between the WH-clause and the pivot in such constructions is similar in essential respects to the question-answer relation in (53) and (56). In both cases, the polarity item is licensed within a full finite clause that is subject to optional elision. Important support for this proposal derives from the generalization, which they establish, that bare NPI pivots are possible in pseudoclefts just in case the full clausal option is also possible (as illustrated in 55).

Weir (2014; 2015) builds on these insights to develop an analysis that is consistent with the idea that fragment answers are in general derived by movement followed by elision of all but the moved element. It also accounts for the marginal character of (53) by assuming that the movement in question takes place in the derivation of phonological forms (and its effects are therefore invisible to semantic conditions such as those essential to the licensing of polarity items) and is a marked and 'last resort' operation – one that is specific to that context and one which applies only to head off the possibility that a focus marked expression might be elided.

These issues are difficult and important, but they clearly do not challenge the empirical claim that was the starting point for the present chapter: that Irish and English differ fundamentally with respect to the well-formedness of NPI fragments. The Irish cases we have been concerned with are in no way marked or recondite; nor are they variably acceptable or restricted to very particular contexts, as is the English possibility in (53). They are simply routine aspects of the grammar of the language: productive and unmarked.

The proposals developed here account for that property of Irish and are also compatible with the framework developed for the English cases by Weir and his predecessors. The difference between the two languages is that the grammar of Irish includes a syntactic operation that is routine and productive, and that renders appeal to marked or 'last resort' operations unnecessary. The grammar of English, by contrast, includes no such operation and in providing a structural description for (53) must rely on the logic of last resort.

4.5 The Syntax of Negation and the Syntax of Narrative Fronting

For the announced purpose of this chapter (resolving the apparent anomaly of NPI fragments), this is arguably as far as we need to go. It has been shown that the syntax of Irish must include a mechanism by which NPIs can be raised to a clause-peripheral position and that in that position raised expressions remain within the semantic scope of sentential negation and NPIs are therefore licensed. That much is close to incontrovertible. In addition, we need to appeal to an ellipsis process that elides the clause from which the NPI has been raised; but the process we must appeal to is of a kind that is familiar, well-studied, and well-attested.

There remain, though, obvious and large questions about how such a system can be integrated within some credible larger theory of Irish clausal syntax. I want to say something about those questions here, but the discussion will of necessity be somewhat unsatisfying because the issues that arise cannot be fully addressed within the scope of a chapter such as this.

Consider first the extended projection of the verb-initial finite clauses of the language. Sentential negation is expressed overtly in these clauses on C; every candidate complementizer in the language (interrogative, WH, conditional, root, default …) is paired with a negative counterpart with which it competes in the system of exponence (see McCloskey 2001 for data and discussion).

However, McCloskey (2017) argues that the expression of negation in finite clauses in Irish is, in fact, distributed over two positions linked by an agreement relation: (i) on C, the highest element of the extended clausal projection where its morphological exponence appears, and (ii) on a lower polarity head, one which appears in an arguably more expected position below the expression of clausal force and that is the position relevant for the semantics of polarity.

In finite clauses, the two polarity positions are separated by a head that has as its exponents the various preverbal tense markers of the language and that expresses a (limited) combination of tense and modality properties. We will call that element here TM1. The finite clause in (57), then, has the schematic structure in Figure 1.

(57) Ní -or óladar aon nimh sa teach.
C -PAST drink.PAST.P3 any poison in.the house
'They did not drink any poison in the house.'

The complement of that head is, in turn, projected by a lower tense-modality head, one which expresses a set of tense and modality distinctions that further refine those introduced by TM1. This head (TM2) is realized morphologically in the

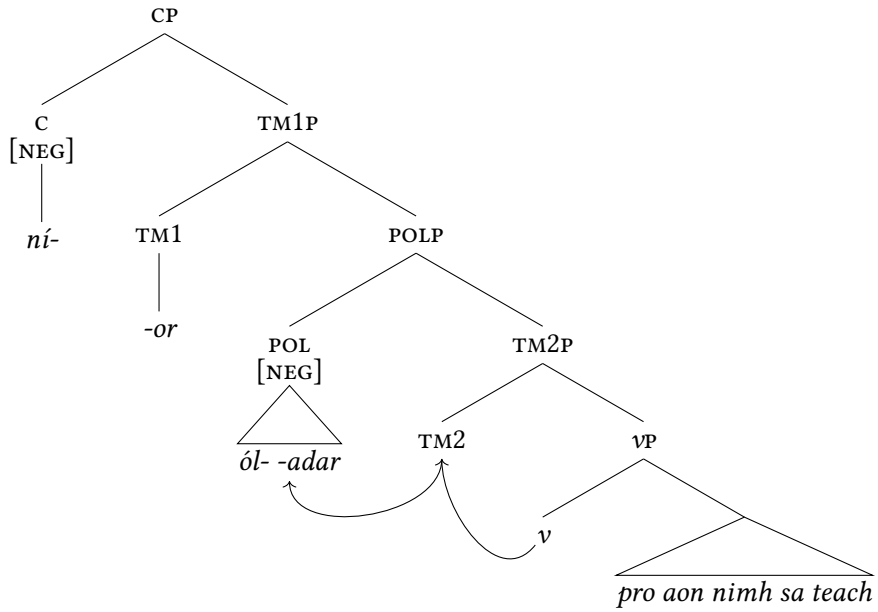


Figure 1: Tree for (57)

post-verbal inflectional suffixes and hosts a raised subject in its specifier, yielding VSO order.

In nonfinite clauses, which are characterized by an absence of head-movement, the lower polarity head hosts the overt marker of negation (*gan*), which clearly appears in a position lower than C, and which precedes subject-position:

(58) Ba mhian leat gan mé creidbheáil ins an rud.
COP.PAST desire with.you NEG.NONFIN me believe.VN in the thing
'You wanted me not to believe in the thing.' UMI 167

How might Narrative Fronting be integrated into this framework? The point reached earlier is summarized in (59), repeated from (41) above.

(59) NARRATIVE FRONTING:
The finite negation head may include a probe that attracts XPs that are alternative-evoking.

We can begin by assuming that there is a feature ALT that identifies alternative-evoking expressions to the syntax. This feature, I assume, is part of the lexical specification of NPIs, one of the ways in which the pure numeral use of *aon*,

for instance, is distinguished from its NPI use. Narrative Fronting, then, will be understood as a probe-goal interaction between a negative head and a phrase in its local domain which bears the ALT-feature; in that interaction, the target phrase is raised to the specifier position of the polarity head.

But given the structure in Figure 1, we now have two candidate heads to consider when asking which head hosts the crucial probe: the higher negative C or the lower polarity head. Given that in the overt structure, the raised XP appears to the left of the negative complementizer, the answer to that question would seem to be obvious. One might think that the ALT-probe is an optional subpart of the negative complementizer and that Narrative Fronting is, therefore, movement to its specifier position.

Alternatively, it could be that the probe that attracts the alternative-evoking expression into its specifier position is hosted by the lower (negative) polarity head. Following raising of the finite verb to the polarity position, the fronted XP will appear to the left of the inflected verb and niched between the two (linked) expressions of negation. (60) would then have the syntax shown in Figure 2.

(60) Focal amháin féin ní-or labhair ceachtar acu.
word one even C.NEG-PAST speak.PAST either at.them
'Not even a single word did either of them speak.' INIT 164

The first option (raising to the specifier of negative C) accounts immediately for the observed order of constituents: the raised XP in Narrative Fronting always precedes the expression of finite negation. But it is problematic in important ways as well. In particular, this analysis defines Narrative Fronting as a movement to the clause-edge, thus risking the incorrect prediction that the operation would have the same locality-profile as the much-studied Ā-movements of the language. It does not, though. As shown in McCloskey (1996), Narrative Fronting is movement to a TP-internal position and is strictly clause-bound.

In addition, the observed word-order in cases like (60) is guaranteed by the proposal developed in McCloskey (1996) that finite complementizers in general in Irish attain their pronounced positions by way of a postsyntactic operation that lowers them, across intervening material if present, to the position of the inflected verb. On this view, C-lowering is an exact analog of English T-lowering (the 'affix-hopping' of Chomsky 1957) and is justified by exactly the same kind of argument. In the case of Irish, C-lowering is one of a set of post-syntactic operations which jointly create the complex morphological word known in the Irish linguistic tradition as the 'verbal complex', whose properties and intricate internal structure have been studied, for example, by Jason Ostrove (2018). In the

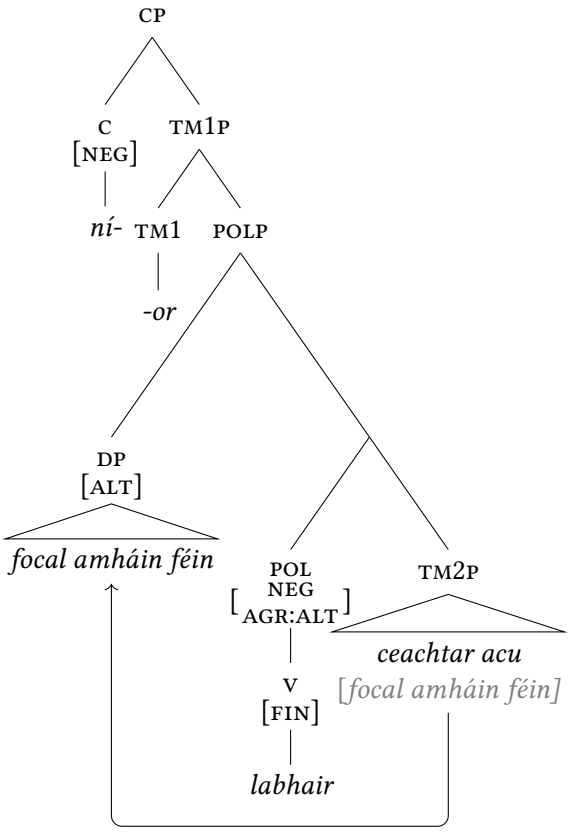


Figure 2: Tree for (60)

case of (60), the effect is to create a complex morphological word of the form in (61) in the position of the polarity head:

(61) { C-NEG⁀TM1⁀ROOT⁀VCE⁀TM2 }

In the context of the c-lowering proposal, the analysis in Figure 2 becomes natural. On this account, the fronted XP is commanded by an expression of negation at every point in its derivational career and standard (and natural) mechanisms provide an understanding of the fact that Narrative Fronting applies only in negative clauses (an important advance over the account offered in McCloskey 1996), while avoiding the risk of predicting that the operation would have the locality-profile of an Ā-movement.

The proposal of c-lowering has been controversial (see for instance Maki & Ó Baoill 2017, who show that one of the strands of evidence offered in McCloskey

1996 was based on a misinterpretation of the relevant evidence), but I know of no competitor proposal that deals with the range of facts it accounts for. In addition, there is no reason for embarrassment in a contemporary theoretical setting concerning appeal to a post-syntactic (as opposed to syntactic) lowering.

In the context of Figure 2, the ellipsis process appealed to in accounting for the fragment answers of, for instance, (30), (31) or (47b) – one which will eliminate all but the phrase fronted by Narrative Fronting – will emerge as a familiar kind of polarity ellipsis triggered by an ellipsis-licensing feature (Merchant's E) on the lower polarity head.[20] Further, following Benjabi & Pesetsky (2022), I will assume that in this case the E-feature extends to the immediate projection of the polarity head, ensuring elimination from the pronounced string of all but the specifier of the licensing head.

Much hard work clearly remains to be done on all of these questions. But I hope to have shown in this brief discussion that the project of integrating the proposals of this chapter into a credible larger theory of clause structure (in Irish) is far from being a hopeless one.

5 Implications

The proposal outlined here (and anticipated in Merchant's paper) depends on a central element of the 'move and delete' approach articulated by Merchant (2004) for subsentential fragments of propositional type. That approach links the ellipsis possibilities found in a given language with the inventory of movement-types available in that language. That inventory in turn reflects the inventory of probes in the language; so we have yet another case in which variation among languages has its roots in combinatorial properties of elements of the functional vocabulary. In the absence of such an approach, it is not clear how to make the required connection between the possibility of fragment NPIs and the existence and properties of Narrative Fronting.

The analysis also preserves the essence of the generalization that NPIs may not function as fragment answers. The possibility of NPI fragments will emerge in a language only if it has a movement operation that is not scope-enhancing and so does not raise the moved item out of the crucial licensing context. Irish does not show that the initial generalization is wrong, then, but rather suggests a refinement.[21] This clarification too depends in a fundamental way on the move-and-delete approach to this kind of fragment response.

[20] For an interesting comparison see especially Gribanova (2017) on polarity ellipses in Russian.

[21] The implication is, obviously, that in diagnosing some constituent as either an NPI or a negative concord item, one should not use the fragment answer diagnostic in a simplistic way. Before

Most important, perhaps, the arguments supporting an ellipsis-based analysis of at least this type of fragment response are quite powerful. Few expression-types are as subtly dependent on context as are NPIs; their interpretation and their well-formedness depend on very particular properties of the compositional settings in which they appear. And even those who have pushed hardest to construct a fundamentally semantic or pragmatic understanding of how NPIs are licensed and interpreted recognize that the definition of that compositional setting has a syntactic component (see, for example Ladusaw 1979: 206–207). Gajewski (2005), Homer (2011), and Homer (2021), in particular, have developed persuasive arguments that even the fundamental semantic constraint – that NPIs must appear in (Strawson) downward-entailing environments – must be stated in terms both syntactic and semantic, as in (62):

(62) An NPI α is licensed in sentence S only if there is a (syntactic) constituent A of S containing α such that A is downward-entailing with respect to the position of α. (Homer 2021: 5)

That is, syntactic constituents (rather than 'operators') are downward or upward entailing (in virtue, of course, of their interpretive properties). In turn, what this means is that the minimizer fragment *smid* in the exchange in (47b) above, if it is to be appropriately interpreted and licensed, must appear within a syntactic constituent whose semantic properties are such that it is (Strawson) downward entailing with respect to the position of *smid*. This is guaranteed in a straightforward way by the ellipsis analysis, because its principal commitment is exactly that the minimizer in (47b) is contained within an interpreted syntactic structure of the required kind – though one that happens not to be pronounced. It is difficult to imagine how the required compositional scaffolding (semantic and syntactic) might be supplied in an approach to fragments that eschewed such silent structure.

But if appeal to ellipsis is correct for this material, there are also implications for what the theory of ellipsis must then look like. Consider again some examples of NPI fragment responses:

(63) a. "Beidh na girseachaí leat." "Ó go bráthach!"
be.FUT the girls with.you ever
'"The girls will be with you." "Oh, never!"' ATFS 505

concluding that an element is a negative concord item on the basis of the test, one needs to ask if the language has an independently available mechanism that could displace an NPI out of a potential ellipsis-site, yielding a pattern like the one discussed here for Irish.

b. “Cá mhéad a bhéas ar sin?” “Pingin ar bith,” arsa bean a’ tsiopa.
what amount C be.PRES on that penny any said woman the shop.GEN
‘“How much will that be?” “Not a penny,” said the shop-woman.’ BRD 81

c. Cé a d’ inis duitse go bhfuil tú ag déanamh mar is ceart? Aon duine.
who C.WH PAST- tell.PAST to.you C be.PRES you PROG do.VN as is right one person
‘Who told you that you were doing the right thing?’ ‘Nobody.’ ISNB 139

d. “Beidh dornán maith mónadh de dhíobháil ortha le teinte a choinneáil ann.” “Fód ar bith.”
be.FUT quantity good peat.GEN of need on.them to fires VCE keep.VN in.it sod any
‘“They’ll need a lot of turf (peat) to keep the fires going in it.” “Not a sod.”’ ST 206

For each example in (63), there is a very natural paraphrase in terms of Narrative Fronting:

(64) a. Go bráthach [ní bheidh na girseachaí liom].
‘Never will the girls be with me.’

b. Pingin ar bith [ní bheidh ar sin].
‘That won’t cost a penny.’

c. Aon duine [níor inis domh go raibh mé ag déanamh mar is ceart].
‘Not one person told me that I was doing the right thing.’

d. Fód ar bith [ní bheidh de dhiobháil orthu].
‘Not one sod will they need.’

In each case, the elided clause must include sentential negation to license the stranded polarity item and to guarantee the right interpretation. However, no matching negation appears in the apparent antecedent. Such examples are very frequent (see also 2, all four in 30, and two in 31 above) and they add to the steady accumulation of evidence in recent years that elided clauses are not required to match their antecedents in polarity (Yoshida 2010; Toosarvandani 2013; Kroll & Rudin 2018; Rudin 2019; Kroll 2019; Anand et al. 2021; Ranero 2021). Further, these

examples are of a particular type: negation appears in the elided clause but not in its antecedent. This is worth noting because in work on sluicing in English it has been observed (Anand et al. 2021) that this pattern is rarer than its inverse (negation in the antecedent, no negation in the ellipsis-site).

The issues around antecedence do not end here though. In question-response pairs, it is usually easy to identify an appropriate overt antecedent. However, NPI fragments also occur in contexts in which there is no such antecedent:

(65) a. Bhain sé triail as an uimhir fóin. Freagra ar bith.
took.PAST he try out.of the number phone.GEN answer any
'He tried the phone number. No answer.' TAIR 114

b. Chuaigh mé amach a dh'éisteacht le ceol na n-éan.
go.PAST I out to-listen with music the.GEN birds.GEN
Fuaim dá laghad.
sound of.the least
'I went out to listen to the singing of the birds. Not a sound.' RNG 20-10-18

c. Níl aon uaigh ann, a Dheaid. Chuartaigh muid an áit
is.not one grave in.it VOC-PTC Dad search.PAST we the place
ó bhun go barr. Tada.
from bottom to top anything
'There's no grave, Dad. We searched the place from top to bottom. Nothing.' A 98

These are all attested examples. (65a) was also checked with six native speaker consultants, all of whom accepted it without hesitation as natural, well-formed, and clear. The crucial property of such cases is that, although the evidence for ellipsis is as clear and as strong as for the other instances of NPI fragments, there is no overt antecedent in the discourse context. However, there is in each case a strongly salient but implicit question (a QUD in the sense of Roberts 2012) – *Would he get an answer? What would he hear? What did they find?* – in the local discourse context.

Given how strong the evidence for ellipsis is in such fragment answer cases, it seems we must conclude (with Merchant 2004: Section 5) that, for certain kinds of ellipsis at least, antecedents with the necessary kind of syntactic and semantic properties may be found in discourse representations rather than in any overt linguistic signal. This conclusion runs against the grain of certain trends in the study of ellipsis, some of which tend to assume that if there is no overt

antecedent there can be no "syntax in the silence" (to use Merchant's phrase). But the conclusion is, in a certain sense, profoundly unsurprising. The initiators of one of the most sophisticated and influential frameworks for the study of discourse dynamics Farkas & Bruce (2010) take the following position:

> We follow the literature … in having a discourse component that records the questions under discussion … and assume that the items on it are syntactic objects paired with their denotations. (Farkas & Bruce 2010: 86)

The reason that they adopt this position is exactly that "the grammar of cross-turn conversation and ellipsis has to have access to the grammatical form (and not just the content) of immediately previous utterances" (Farkas & Bruce 2010: 86).

6 Conclusion

Working on ellipsis is very difficult and can feel frustrating. The problems are unimaginably more challenging than they seemed to be fifty years ago and it is easy to yield sometimes to the feeling that little theoretical progress has been made in the interim, even though there are uncountably many more observations available to us now that might shape theory-development.

Part of the difficulty is that it remains unclear for what phenomena the general theory of ellipsis is supposed to be responsible. The case of subsentential fragments is a classic one of that type; the exchange between Stainton (2006) and Merchant (2010) in particular makes that very clear. I take the present chapter to be a modest contribution to that difficult and important area.

However, working on ellipsis also often brings useful spin-offs. In the present case, it has led to a more complete survey of the landscape of polarity-sensitive items in Irish than has been available before. It has also led to a better understanding of Narrative Fronting – one of the more intriguing components of the syntax-pragmatics interface in Irish – than has been available before.

And finally the chapter has also, I hope, provided a clearer view of what the extended projection of finite clauses in Irish looks like. Research on extended projections and research on ellipsis are deeply entangled because the correct theory of the extended projections for a given language must provide a basis for understanding what the inventory of ellipsis processes in that language are.

Acknowledgements

This chapter has its origins in conversations with Jason Merchant some twenty years ago. In bringing it finally to completion I owe an even larger debt of gratitude than usual to the friends and colleagues who have acted as my linguistic guides: Caitlín Nic Niallais, Lillis Ó Laoire, Máire Ní Neachtain, Pádhraic Ó Ciardha, Róise Ní Bhaoill and Seosaimhín Ní Bheaglaoich. I am also grateful for the advice, discussion, and suggestions of Paolo Acquaviva, Pranav Anand, Liam Breatnach, Lisa Cheng, Sandy Chung, Nicola D'Antuono, Vera Gribanova, Bill Ladusaw, Anikó Liptak, Gillian Ramchand, Ivy Sichel and Gary Thoms. Presentations at UC Santa Cruz in November 2019 and at Leiden University in January of 2020 were very helpful in pushing the project along, as were comments and suggestions made by an anonymous reviewer. This research was supported by funding from the National Science Foundation via Award Number 1451819:*The Implicit Content of Sluicing* (principal investigators Pranav Anand, Daniel Hardt and James McCloskey) to the University of California Santa Cruz.

Appendix A The licensing environments

This appendix presents data concerning the range of licensing environments in which the polarity-sensitive expressions discussed in Section 2 may appear.

A.1 In questions

All of the PSEs of Section 2 appear in polar questions, root and embedded:

(66) EMBEDDED POLAR QUESTIONS:

a. féachaint an samhlódh faic dom
to-see-if C.Q imagine.IMPERS anything/nothing to.me
'to see if anything would spring to mind for me' AGMTS 3

b. go bhfeicfeadh muid an raibh tada beo tar éis na hóiche
C see.COND we C.Q be.PAST anything alive after the night
'so that we could see if anything was alive after the night' IMSBRM 55

(67) ROOT POLAR QUESTIONS:

a. An féidir liom tada a fháil duit?
C.Q possible with.me anything get.NFIN for.you
'Can I get anything for you?' DEAD 149

b. Agus a-r ghnóthaigh sí aon rás eile i mbliana?
and C.Q-.PAST win.PAST she one race other this-year
'And did she win any other race this year?" RNG 26-08-19

The issue of under what conditions NPIs are licensed in WH-questions has been important in debates about the nature of NPI-licensing (Giannakidou 1999, 2011, Guerzoni & Sharvit 2007, Mayr 2013). The PSEs we care about here appear in such questions, root and embedded:

(68) WH-QUESTIONS:

a. Cé eile a bhfuil fhios aige tada faoi seo?
who other C be.PRES knowledge at.him anything about this
'Who else knows anything about this?' CAB 89

b. Goidé mar a thiocfadh liom aon fhear a phósadh?
how C come.COND with.me any man VCE marry.VN
'How could I marry any man?' AM 162

c. Cén chaoi a mbeadh fhios acu tada?
what way C be.COND knowledge at.them anything
'How would they know anything?' SMC 304

d. Cé a chreidfeadh go raibh sé ariamh bródúil?
who C believe.COND C be.PAST he ever proud
'Who would believe that he was ever proud?' ATFS 51

Many of these cases involve rhetorical questions, e.g. (68b–d). Others are information-seeking questions which require exhaustive answers, as in (68a), a pattern which is consistent in particular with the work of Guerzoni & Sharvit (2007) and Mayr (2013). Whether or not this pattern holds consistently is something which remains to be investigated.

A.2 In conditional clauses

The PSEs of Section 2 may also appear, with their characteristic existential interpretation, in conditional clauses, both realis (69a) and irrealis (69b,c):

(69) a. Má chaitheann tú choíchín tada a rá, abair
if must.PRES you ever anything say.NFIN say.IMPV
go cuí é.
appropriately it
'If you ever have to say anything, say it appropriately.' AE 62

b. Dá mbeadh baint ar bith leis na gnóithe agam.
if be.COND connection any with the business at.me
'if I had any say in the matter' CLM 38

c. dá dtarlaíodh tada dá hathair
if happen.COND anything to.her father
'if anything should happen to her father' AA 177

A.3 In excessive-degree clauses

Phrases in construction with the degree-word *ró-* ('too') also freely host PSEs:

(70) a. Tá sé ró- mhall anois tada a dhéanamh.
be.PRES it too late now anything VCE do.VN
'It's too late to do anything now.' AA 158

b. bhí sé ró- the chun éinne bheith ag spaisteoireacht
be.PAST it too hot for anyone be.NFIN PROG stroll.VN
'It was too hot for anyone to be out walking.' DPB 102

c. duí gainmhe a bhí ró- aimhréidh d' aon ghalfchúrsa
dune sand.GEN C be.PAST too uneven for any golf-course
'a sand-dune that was too uneven for any golf-course' LGL 170

A.4 In equative, comparative and superlative clauses

(71) EQUATIVE CLAUSES:

a. Bhí cuma air comh folláin agus bhí ariamh
be.PAST appearance on.him as healthy and be.PAST ever
air,
on.him
'He looks as healthy as he ever did.' ATFS 565

b. culaith éadaigh chomh deas agus a chuir aon fhear riamh
suit clothes.GEN as nice as C.WH put.PAST any man ever
ar a dhroim
on his back
'as nice a suit of clothes as any man ever put on his back' EMPP 163

(72) COMPARATIVE CLAUSES:

a. Tá níos mó chéill i gcuid cainnte Dhomhnaill ná
be.PRES more sense.GEN in share talk.GEN than
aidmhigheas aon duine.
admit.PRES.WH any person
'There's more sense in Domhnall's talk than anyone admits.' MO 135

b. Bhí fhios aige-sean níos fearr ná bhí
be.PAST knowledge at.him–CONTR more better than be.PAST
fhios ag duine ar bith. …
knowledge at person any
'He knew better than anyone knew that … ' AM 415

c. Is fusa éinne a smachtú ná do chuid
COP.PRES easy.COMPAR anyone VCE discipline.VN than your portion
féin.
REFL.LOG
'It's easier to discipline anyone than your own (children).' LGL 122

(73) SUPERLATIVE CLAUSES:

a. an obair thógála ba deacra dár déanadh
the work building.GEN COP.PRES difficult.COMPAR C do.PAST.IMPERS
go hiomlán as Gaeilge ariamh
entirely out.of Irish ever
'the most difficult building-work that was ever done entirely through Irish' CCT 199

b. an mheancóg a ba mhó a rinne mé
the mistake C.WH COP.PAST big.COMPAR C.WH make.PAST I
ariamh
ever
'the biggest mistake that I ever made' ATFS 127

See Hoeksema (1983), von Fintel (1999), Gajewski (2010), Herdan & Sharvit (2006), Bumford & Sharvit (2022), and Howard (2014).

I will include in this group relative clauses attached to head nouns modified by the ordinal *céad* ('first'), though the status of such phrases as superlatives remains controversial. PSEs are licensed here:

(74) a. Siod é an chéad am a labhair mé air le aon duine
that it the first time C speak.PAST I on.it with any person
'That was the first time that I spoke about it with anyone.' SMC 287

b. an chéad ál a bhí ariamh aici
the first litter C be.PAST ever at.her
'the first litter she ever had' CA 26

A.5 In certain temporal clauses

As in many other languages, PSEs in Irish appear within certain temporal clauses, notably those introduced by 'before' or 'when', but not those introduced by 'after' (Linebarger 1987; Condoravdi 2010; Krifka 2010).

(75) a. sul a raibh am ag an sáirsint aon cheo a rá leis
before C be.PAST time at the sergeant one mist VCE say.VN with.him
'before the sergeant had time to say the slightest thing to him' LSC 202

b. sul má bheidh a fhios aige tada
before C be.FUT knowledge at.him anything
'before he knows anything' CF 129

c. sul má lonnaigh aon duine ariamh ann
before C settle.PAST any person ever there
'before anyone ever settled there' MABAT 74

(76) a. Bhí siad go han-mhaith dó nuair a bhí a dhath
be.PAST they PTC very-good to.him when C be.PAST anything
acu.
at.them
'They were very good to him when they had anything.' GOG 266

b. nuair a theastaíodh dada uaithi bhuaileadh sí
when C need.PAST.HABIT anything from.her hit.PAST.HABIT she
cnag ar an urlár
knock on the floor
'When she needed anything, she would bang on the floor.' CG 20

c. Ní tostach dóibh nuair a thagas aon bhac in a
COP.NEG silent to.them when C come.PRES any obstacle in their
mbealach.
way
'They are not silent when any obstacle gets in their way.' CG 34

A.6 In arguments of 'negative' predicates

The PSEs of Section 2 also appear in the complements of certain predicates whose meaning has a 'negative' component (in a sense that remains to be clarified). For instance, as seen in (77), they appear in the complements of so-called adversative attitude predicates – those which give rise to an implicature that the holder of the attitude has a negative view of the semantic content of that complement.

(77) a. Bhí imní orainn go dtarlódh aon cheo dhuit.
be.PAST worry on.us C happen.COND any mist to.you
'We were worried that anything would happen to you.' CGC 68

b. Is deacair tada a rá.
COP hard anything say.NFIN
'It's hard to say anything.' FB 146

c. ar eagla go dtarlódh a dhath dó
on fear C happen.COND anything to.him
'for fear that anything might happen to him' TAIR 182

d. Is mór a' trua go gcaithfidh aon duine imeacht as.
COP.PRES great the pity C must any person leave.NFIN out-of.it
'It's a great pity that anyone has to leave it.' BRD 246

I include in this class some implicative predicates like *fail*:

(78) a. Chinn orm éinne a aimsiú.
fail.PAST on.me anyone VCE find.VN
'I failed to find anyone.' TUAIR 04-10-21

b. Chinn air tada in-ite a fháil.
fail.PAST on.him anything edible VCE find.VN
'He failed to find anything edible.' ATT 101

Here, though, there is an interesting contrast between Irish and English. In Irish (and also in Korean, judging by Lee 1995), PSEs may appear as direct arguments of the relevant predicates (rather than appearing only within their complements):

(79) Chinn sé ar aon dochtúir thall ann é a bhaint amach.
fail.PAST it on any doctor over there it VCE take.VN out
'No doctor over there succeeded in extracting it.' SJCCF 275
(*lit.*: 'Any doctor over there failed to extract it.')

An interesting subclass of such predicates are those which express reluctance on the part of an experiencer. These are among Karttunen's (1971) negative implicatives; "they seem to incorporate negation" (Karttunen 1971: 352) and license the inference that in the preference-worlds of their experiencers the eventuality described in their complement is not instantiated.[22] NPIs in English are licensed within the complements of such predicates, but not in the experiencer argument-position. In Irish they are, in addition, licensed in the experiencer argument-position. I emphasize the contrast between the two languages in this respect by offering for the examples in (80) literal English translations, which are ill-formed but, interestingly, understandable.[23]

(80) a. Bhí drogall ar aon imreoir suí leis ag clár
be.PAST reluctance on any player sit.NFIN with.him at board
na himeartha.
the.GEN playing.GEN
'Any player was reluctant to sit with him at the gaming-table.'
SJCCF 337

b. Bhí leisce ar aon duine tada a rá.
be.PAST reluctance on any person anything VCE say.VN
'Anyone was reluctant to say anything.' LGN 96

Similarly, the predicate *cuma* ('matterless, insignificant') allows NPIs in its two argument-positions:

(81) a. Ba chuma liom faoi thada.
COP.PAST matterless with.me about anything
'I didn't care about anything.' RNG 31-01-20

b. Is cuma le ceachtar agaibh fá'n duine eile.
COP.PRES matterless with either of.you about.the person other
'Neither of you cares about the other.' ATFS 425
(*lit.* 'Either of you doesn't care about the other.')

The English facts here may be complicated by a language-particular oddity – a requirement that the licensing element precede the licensed NPI (Ladusaw 1979:

[22] For a sophisticated and relevant discussion, see von Fintel (1999: 115–121).

[23] One might avoid this puzzle by taking the examples in (79), (80) and (81) to involve "free choice" readings. But this move is unlikely to be tenable given that the licensing conditions for such readings are not satisfied in such cases. In addition, the possibilities shown in (79–81) hold only for predicates whose meaning in some sense includes a negative component (Klima 1964, Karttunen 1971, Kadmon & Landman 1993, von Fintel 1999).

206–207). It is also possible that the relevant arguments in Irish are internal and therefore within the scope of the (negative component of the) licensing predicate.

A.7 In the restrictive clause of a universal quantification structure

The final licensing environment to be documented is once again a familiar one: the restrictive clause of a universal quantification structure.

The most frequent case of this type involves NPIs which have "free choice" or quasi-universal readings. On this possibility (in English) see Horn (2000: 163). A relative clause attached to such an element acts as the restrictor for the universal quantification that it expresses and PSEs appear freely.

(82) a. éinne go bhfuil aon chiall aige
anyone C be.PRES any sense at.him
'anyone who has any sense' G 11

b. fear ar bith a bhfuil fhios aige tada
man any C be.PRES knowledge at.him anything
'any man who knows anything' CNF 25

c. am ar bith a mbeadh a dhath le rádh aige liom
time any C be.COND anything to say.VN at.him with.me
'any time he had anything to say to me' LCS 201

Unconditional clauses, which in Irish are marked by means of a distinctive WH-determiner (*pé*, or *cibé*), similarly host PSEs:

(83) pé pingin a bhí riamh acu
whatever penny C be.PAST ever at.them
'whatever pennies they ever had' AI 199

It is hardly a surprise, then, that PSEs are also licensed within relative clauses that restrict nominals headed by explicit universal quantifiers:

(84) a. timpeall ar chuile áit a gceapfaidís a mbeadh aon deis
around on every place C think.COND.P3 C be.COND any opportunity
ag an ngail a bheith ag éalú
at the steam be.NFIN PROG escape.VN
'around every place that they'd think there was any opportunity for the steam to escape'

b. Gach neach a thug cath éagórach riamh d' Fhionn
every being C give.PAST battle unjust ever to
'every being who ever joined battle unjustly with Fionn' SNAF 226

Appendix B Sources of examples

A: *Aileach*, Jackie Mac Donncha, Cló Iar-Chonnacht, 2010
AA: *Athaoibhneas*, Pádhraic Óg Ó Conaire, Sáirséal agus Dill, 1959
AAG: *As an nGéibheann*, Máirtín Ó Cadhain, Sáirséal agus Dill, 1973
ABFS: *An Baile i bhFad Siar*, Domhnall Mac an tSíthigh, Coiscéim, 2000
ABHM: *Ábhar Machnaimh*, An tAthair Donncha Ó Corcora, Foilseacháin Náisiúnta Teoranta, 1985
ACED: *An Chuid Eile Díom Féin – Aistí le Máirtín Ó Direáin*, ed. Síobhra Aiken, Cló Iar-Chonnacht, 2018
AE: *An Eochair*, Máirtín Ó Cadhain, Dalkey Archive Press, 2015
AFAP: *An Fear a Phléasc*, Mícheál Ó Conghaile, Cló Iar-Chonnachta, 1997
AG: *An Gabhar Sa Teampall*, Mícheál Ua Ciarmhaic, Coiscéim, 1986
AGFC: *An Grá Faoi Cheilt*, Pádraig Ó Cíobháin, Coiscéim, 1992
AGMTS: *Ar Gach Maoilinn Tá Síocháin*, Pádraig Ó Cíobháin, Coiscéim, 1991
AG: *An Gabhar Sa Teampall*, Mícheál Ua Ciarmhaic, Coiscéim, 1986
AI: *Allagar na hInise*, Tomás Ó Criomhthain, Oifig an tSoláthair, 1977
AL: *Abair Leat*, Joe Daly, ed. Pádraig Tyers, An Sagart, 1999
AM: *An Mhiorbhailt*, C.B Kelland, trans. Niall Mac Suibhne, Oifig Díolta Foillseacháin Rialtais, 1936
AN: *Athnuachan*, Máirtín Ó Cadhain, Coiscéim, 1995
ANNF: *Ar Nós na bhFáinleog*, Siobhán Ní Shúilleabháin, Coiscéim, 2004
APB: *An Prionsa Beag*, Antoine de Saint-Exupéry, trans. Eoghan Mac Giolla Bhríde, Éabhlóid, 2015
ATFS: *Ag Teacht Fríd an tSeagal*, Helen Mathers, trans. Seosamh Mac Grianna, An Gúm, 1932
ATIM: *An tIolrach Mór, Díoghluim Gearr-Sgéal*, Pádhraic Ó Domhnalláin, Brún agus Ó Nualláin, 1941
ATT: *An tSraith Tógtha*, Máirtín Ó Cadhain, Sáirséal agus Dill, 1977
BM: *Bullaí Mhártain*, Síle Ní Chéileachair agus Donncha Ó Céilleachair, Sáirséal agus Dill, 1969
BRD: *Bean Ruadh de Dhálach*, Séamus Ó Grianna, Oifig Díolta Foillseacháin Rialtais, 1966
BTFS: *An Blascaod Trí Fhuinneog na Scoile*, Nóra Ní Shéaghdha, ed. Pádraig Ó hÉalaí, An Sagart, 2015
C: *Clochmhóin*, Joe Steve Ó Neachtain, Cló Iar-Chonnachta, 1998
CA: *Cnuasacht Airneáin*, Colm Ó Ceallaigh, Coiscéim, 2006
CAA: *Conamara agus Árainn 1880–1890: Gnéithe den Stair Shóisialta*, Micheál Ó Conghaile, Cló Iar-Chonnachta, 1988

CAB: *Carraig an Bháis*, Colm Ó Ceallaigh, Coiscéim, 2007
CC: *Cruithneacht agus Ceannabháin*, Tomás Bairéad, Comhlucht Oideachais na hÉireann, 1940
CCT: *Camchuairt Chonamara Theas*, Tim Robinson, trans. Liam Mac Con Iomaire, Coiscéim, 2002
CDC: *Castar na Daoine ar a Chéile, Scríbhinní Mháire 1*, Séamus Ó Grianna, ed. Nollaig Mac Congail, Coiscéim, 2002
CF: *Cois Fharraige Le Mo Linnse*, Seán Ó Conghaile, Clódhanna Teoranta, 1974
CFC: *Céad Fáilte go Cléire*, ed. Marion Gunn, An Clóchomhar Tta, 1990
CG: *Ceol na nGiolcach*, Padhraic Óg Ó Conaire, Oifig an tSoláthair, 1968
CGC: *Caillte i gConamara, Scéalta Aniar*, ed. Brian Ó Conchubhair, Cló Iar-Chonnacht, 2014
CHD: *Chicago Driver*, Maidhc Dainín Ó Sé, Coiscéim, 1992
CLM: *Cúl le Muir agus Scéalta Eile*, Séamus Ó Grianna, Oifig an tSoláthair, 1961
CNF: *Clann na Feannóige*, Colm Ó Ceallaigh, Coiscéim, 2004
COC: *Cora Cinniúna*, Séamus Ó Grianna, ed. Niall Ó Domhnaill, An Gúm, 1993
CTP: *Cuimhne an tSeanpháiste*, Micheál Breatnach, Oifig an tSoláthair, 1966
D: *An Draoidín*, Séamus Ó Grianna, Oifig Díolta Foillseacháin Rialtais, 1959
DEAD: *Déirc an Díomhaointis*, Pádhraic Óg Ó Conaire, Sáirséal agus Dill, 1972
DGD: *Deoir Ghoirt an Deoraí*, Colm Ó Ceallaigh, Cló Iar-Chonnachta, 1993
DPB: *Dualgas Pheadair Bhig*, trans Séamus Ó Maolchathaigh, Oifig an tSoláthair, 1953
DR: *Dracula*, Bram Stoker, trans. Seán Ó Cuirrín, An Gúm, 1933/1997
EMPP: *Eachtraí Mara Phaidí Pheadair as Toraigh*, Séamus Mac a' Bhaird, ed. Aingeal Nic a' Bhaird, Caoimhín Mac a' Bhaird, Nollaig Mac Congail, Arlen House, 2019
FB: *Feamain Bhealtaine*, Máirtín Ó Direáin, An Clóchomhar Tta., 1961
FF: *Fonn na Fola*, Beairtle Ó Conaire, Cló Iar-Chonnacht, 2005
G: *Greenhorn*, Maidhc Dainín Ó Sé, Coiscéim, 1997
GDDR: *Go dTaga do Ríocht, Boicíní Bhóthar Kilburn, Cripil Inis Meáin*, Mícheál Ó Conghaile, Cló Iarr-Chonnachta, 2009
GFH: *An Ghlan-fhírinne*, Cóil Learaí Ó Finneadha, Cló Iarr-Chonnacht, 2014
GOG: *Glórtha ón Ghorta: Béaloideas na Gaeilge agus an Gorta Mór*, Cathal Póirtéir, Coiscéim, 1996
GSA: *An Giorria San Aer*, Ger Ó Cíobháin, Coiscéim, 1992
IAE: *In Aimsir Emmet*, James Murphy trans. Colm Ó Gaora, Oifig Díolta Foillseacháin Rialtais, Dublin, 1937

IMSBRM: *Idir Mná: Scríbhneoirí Ban Ros Muc*, ed. Máire Holmes, Pléaráca Chonamara, 1995
INIT: *Idir Neamh is Talamh*, Joe Steve Ó Neachtain, Cló Iar-Chonnacht, 2014
ISNB: *Iad Seo Nach Bhfaca*, Beairtle Ó Conaire, Cló Iar-Chonnacht, 2010
LAN: *Leoithne Aniar*, ed. Pádraig Tyers, Cló Dhuibhne, Baile an Fhéirtéaraigh, 1982
LCS: *Le Clap-Sholus*, Séamas Ó Grianna, Oifig an tSoláthair, 1967
LGL: *Le Gealaigh*, Pádraig Ó Cíobháin, Coiscéim, 1991
LGN: *Le Gean agus scéalta eile*, Mike P. Ó Conaola, Sián, 2020
LL: *Lámh Láidir*, Joe Steve Ó Neachtain, Cló Iar-Chonnachta, 2005
LOFRS: *Liam Ó Flaithearta - Rogha Scéalta*, trans. Micheál Ó Conghaile, Cló Iar-Chonnacht, 2020
LSC: *Lig Sinn i gCathú*, Breandán Ó hEithir, Sáirséal agus Dill, 1976
MABAT: *Mar a Bhí Ar dTús: Cuimhne Seanghasúir*, Joe Steve Ó Neachtain, Cló Iarr-Chonnacht, 2018
MO: *Muintir An Oileáin*, Peadar O'Donnell, trans. Seosamh Mac Grianna, Oifig Díolta Foillseacháin Rialtais, 1952
NLAB: *Na Laetha a Bhí*, Eoghan Ó Domhnaill, Oifig an tSoláthair, 1968
OMGS: *Ó Mhuir go Sliabh*, Séamus Ó Grianna, Oifig an tSoláthair, 1961
OOGC: *Ó Oileán go Cuilleán*, Muiris Ó Súilleabháin, ed. Nuala Uí Aimhirgín, Coiscéim, 2000
OTA: *Ón tSeanam Anall, Scéalta Mhicí Bháin Uí Bheirn*, ed. Mícheál Mac Giolla Easbuic, Cló Iar-Chonnachta, 2008
PAA: *Peacaí Ár nAthaireacha*, Mícheál Ó Súilleabháin, Coiscéim, 1992
PI: *Punt Isló*, Maidhc Dainín Ó Sé, Coiscéim, 2013
PNG: *Pobal na Gaeltachta*, ed. Gearóid Ó Tuathaigh, Liam Lillis Ó Laoire, Seán Ua Súillebháin, Cló Iar-Chonnachta, 2000
RNG: *Raidió na Gaeltachta* (numerical index refers to date of broadcast)
SIF: *Seanchas Iascaireachta agus Farraige*, Seán Ó hEochaidh, *Béaloideas* 33, 1965
SJCCF: *Seanchas Jimmí Chearra Chois Fharraige*, ed. Pádraic Ó Cearra, Coiscéim, 2010
SJSJ: *Seachrán Jeaic Sheáin Johnny*, Mícheál Ó Conghaile, Cló Iarr-Chonnacht, 2002
SK: *Sáile Chaomhánach*, C.J. Kickham, trans. Máirtín Ó Cadhain, An Gúm, 1932/1986
SMC: *Stairsheanchas Mhicil Chonraí – Ón Máimín go Ráth Chairn*, ed. Conchúr Ó Giollagáin, Cló Iar-Chonnachta, 1999
SNAF: *Seanchas na Féinne*, Niall Ó Dónaill, An Gúm, 1943/1996

SRNF: *Seanchas Rann na Feirste*, Maelsheachlainn Mac Cionaoith, 2006
SSOTC: *Síscéalta ó Thír Chonaill*, ed. Seán Ó Heochaidh, Máire Ní Néill and Séamas Ó Catháin, Comhairle Bhéaloideas Éireann, 1977
ST: *An Sean-Teach*, Séamas Ó Grianna, Oifig an tSoláthair, 1968
TAIR: *Tairngreacht*, Proinsias Mac a' Bhaird, Leabhair Comhar, 2018
TII: *Tone Inné agus Inniu*, Máirtín Ó Cadhain, Coiscéim, 1992
TMGB: *Thiar sa Mhainistir atá an Ghaolainn Bhreá*, Brighid Ní Mhóráin, An Sagart, 1997
TUAIR: *Tuairisc*: online newspaper: https://tuairisc.ie (numerical index refers to date of publication)
UMI: *Uaill-Mhian Iúdaigh*, Roy Bridges, trans. Tadhg Ó Rabhartaigh, Oifig Díolta Foillseacháin Rialtais, 1936

References

Acquaviva, Paolo. 1996. Negation in Irish and the representation of monotone decreasing quantifiers. In Robert D. Borsley & Ian Roberts (eds.), *The syntax of the Celtic languages: A comparative perspective*, 284–317. Cambridge: Cambridge University Press.

Anand, Pranav, Daniel Hardt & James McCloskey. 2021. The Santa Cruz sluicing dataset. *Language* 97(1). e68–e88. DOI: 10.1353/lan.2021.0009.

Barker, Chris. 2018. Negative polarity as scope marking. *Linguistics and Philosophy* 41. 483–510. DOI: 10.1007/s10988-018-9234-2.

Bauer, Winifred. 1997. *The Reed reference grammar of Māori.* Auckland, New Zealand: Reed Publishing.

Benjabi, Ido & David Pesetsky. 2022. E-extension and the uniformity of silence. Presentation in the *You're on Mute* Seminar Series, October 7th 2022.

Bernini, Giuliano & Paolo Ramat. 1996. *Negative sentences in the languages of Europe: A typological approach.* Berlin: de Gruyter.

Bjorkman, Bronwyn & Hedde Zeijlstra. 2019. Checking up on ϕ-agree. *Linguistic Inquiry* 50. 527–569. DOI: 10.1162/ling_a_00319.

Bolinger, Dwight. 1972. *Degree words.* The Hague: Mouton.

Bumford, Dylan & Yael Sharvit. 2022. Negative polarity items in definite superlatives. *Linguistic Inquiry* 53. 255–293. DOI: 10.1162/ling_a_00409.

Chierchia, Gennaro. 2013. *Logic in grammar: Polarity, free choice, and intervention.* Oxford: Oxford University Press.

Chomsky, Noam. 1957. *Syntactic structures* (Janua Linguarum, Nr. 4). The Hague: Mouton & Co.

Condoravdi, Cleo. 2008. Punctual *until* as a scalar NPI. In Kristin Hanson & Sharon Inkelas (eds.), *The nature of the word: Studies in honor of Paul Kiparsky*, 631–653. Cambridge: MIT Press. DOI: 10.7551/mitpress/7894.003.0033.

Condoravdi, Cleo. 2010. NPI licensing in temporal clauses. *Natural Language and Linguistic Theory* 28. 877–910. DOI: 10.1007/s11049-010-9115-z.

Crnič, Luka. 2014. Against a dogma on NPI licensing. In Luka Crnič & Uli Sauerland (eds.), *The art and craft of semantics: A festschrift for Irene Heim.* (MITWPL 70, Volume One), 117–145. Cambridge: Department of Linguistics, MIT.

Crnič, Luka. 2019. *Any*: Logic, likelihood, and context (pts. 1 and 2). *Language and Linguistic Compass* 13. e12350, e12351. DOI: 10.1111/lnc3.12354.

D'Antuono, Nicola. 2023. The syntax of emphatic negation in Modern Irish. *Glossa, a Journal of General Linguistics* 9(1). 1–37. DOI: 10.16995/glossa.10369.

D'Antuono, Nicola. 2024. *An emergent theory of negation.* Università di Padova. (Doctoral dissertation).

Dayal, Veneeta. 1998. *Any* as inherently modal. *Linguistics and Philosophy* 21. 433–476.

Dayal, Veneeta. 2013. A viability constraint on alternatives for free choice. In A. Falaus (ed.), *Alternatives in semantics*, 88–122. London: Palgrave McMillan.

de Swart, Henriëtte. 1996. Meaning and use of *not … until*. *Journal of Semantics* 13. 221–263. DOI: 10.1093/jos/13.3.221.

Deal, Amy Rose. 2022. Negative concord as downward agree. In Özge Bakay, Breanna Pratley, Eva Neu & Peyton Deal (eds.), *NELS 52, Proceedings of the Fifty-Second Meeting of the Northeastern Linguistic Society*, 235–244. Amherst, Massachusetts: University of Massachusetts, Graduate Student Linguistic Association.

Declerck, Renaat. 1995. The problem of *not … until*. *Linguistics* 33. 51–98. DOI: 10.1515/ling.1995.33.1.51.

den Dikken, Marcel, André Meinunger & Chris Wilder. 2000. Pseudoclefts and ellipsis. *Studia Linguistica* 54. 41–89. DOI: 10.1111/1467-9582.00050.

Duffield, Nigel. 1995. *Particles and projections in Irish syntax*. Dordrecht: Kluwer Academic Publishers.

Farkas, Donka & Kim B. Bruce. 2010. On reacting to assertions and polar questions. *Journal of Semantics* 27. 81–118. DOI: 10.1093/jos/ffp010.

Fauconnier, Gilles. 1975. Pragmatic scales and logical structure. *Linguistic Inquiry* 6. 353–375.

Fauconnier, Gilles. 1979. Implication reversal in a natural language. In F. Guenther & S. J. Schmidt (eds.), *Formal semantics and pragmatics for natural languages*, 188–199. Dordrecht: Reidel Academic Publishing.

Gajewski, Jon. 2005. *Neg-raising: Polarity and presupposition.* Massachusetts Institute of Technology. (Doctoral dissertation).

Gajewski, Jon. 2010. Superlatives, NPI's and *most*. Manuscript, University of Connecticut.

Gajewski, Jon. 2011. Licensing strong NPI's. *Natural Language Semantics* 19. 109–148. DOI: 10.1007/s11050-010-9067-1.

Gajewski, Jon & I-ta Chris Hsieh. 2014. Comments on negative polarity items in definite descriptions. In Luka Crnič & Uli Sauerland (eds.), *The art and craft of semantics: A Festschrift for Irene Heim.* (MTIWPL 70, Volume One), 181–198. Cambridge: Department of Linguistics, MIT.

Giannakidou, Anastasia. 1992. *Until*, aspect and negation: A novel argument for two *untils*. In Brendan Jackson (ed.), *SALT 12, Proceedings of the 12th Semantics and Linguistic Theory Conference*, 84–103. San Diego: Linguistic Society of America.

Giannakidou, Anastasia. 1998. *Polarity sensitivity as (non)veridical dependency*. Amsterdam: John Benjamins Publishing Company.

Giannakidou, Anastasia. 1999. Affective dependencies. *Linguistics and Philosophy* 22. 367–421. DOI: 10.1023/A:1005492130684.

Giannakidou, Anastasia. 2000. Negative … concord. *Natural Language and Linguistic Theory* 18. 457–523.

Giannakidou, Anastasia. 2001. The meaning of Free Choice. *Linguistics and Philosophy* 24. 659–735. DOI: 10.1023/A:1012758115458.

Giannakidou, Anastasia. 2011. Negative polarity and positive polarity: Licensing, variation, and compositionality. In Klaus von Heusinger, Claudia Maienborn & Paul Portner (eds.), *The handbook of natural language meaning*, 2nd edition, 1660–1712. Berlin: Mouton de Gruyter.

Giannakidou, Anastasia & Hedde Zeijlstra. 2017. The landscape of negative dependencies: N-words and negative concord. In Martin Everaert & Henk van Riemsdijk (eds.), *The Blackwell companion to syntax, second edition*. Hoboken: John Wiley & Sons.

Gribanova, Vera. 2017. Head movement and ellipsis in the expression of Russian polarity focus. *Natural Language and Linguistic Theory* 35. 1079–1121. DOI: 10.1007/s11049-017-9361-4.

Grice, H. Paul. 1975. The logic of conversation. In Peter Cole & Jerry Morgan (eds.), *Syntax and semantics 3: Speech acts*, 43–58. New York: Academic Press. DOI: 10.1163/9789004368811_003.

Guerzoni, Elena & Yael Sharvit. 2007. A question of strength: On NPI's in interrogative clauses. *Linguistics and Philosophy* 30. 361–391. DOI: 10.1007/s10988-007-9014-x.

Haegeman, Liliane & Raffaella Zanuttini. 1991. Negative heads and the Neg Criterion. *The Linguistic Review* 8. 233–251.

Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Oxford University Press.

Heim, Irene. 1984. A note on negative polarity and downward entailingness. In Catherine Jones & Peter Sells (eds.), *NELS 14, Proceedings of the Fourteenth Meeting of the Northeastern Linguistic Society*, 98–107. Amherst, Massachusetts: University of Massachusetts, Graduate Student Linguistic Association.

Herburger, Elena. 2001. The negative concord puzzle revisited. *Natural Language Semantics* 9(3). 289–333. DOI: 10.1023/A:1014205526722.

Herdan, Simona & Yael Sharvit. 2006. Definite and nondefinite superlatives and NPI licensing. *Syntax* 9. 1–31. DOI: 10.1111/j.1467-9612.2006.00082.x.

Hoeksema, Jack. 1983. Negative polarity and the comparative. *Natural Language and Linguistic Theory* 1. 403–434. DOI: 10.1007/BF00142472.

Hoeksema, Jack. 2000. Negative polarity items: Triggering, scope and c-command. In Laurence R. Horn & Yasuhiko Kato (eds.), *Negation and polarity: Syntactic and semantic perspectives*, 115–146. Oxford: Oxford University Press.

Homer, Vincent. 2011. *Polarity and modality*. University of California, Los Angeles. (Doctoral dissertation).

Homer, Vincent. 2020. Negative polarity. In Lisa Matthewson, Cécile Meier, Hotze Rullman & Thomas Ede Zimmermann (eds.), *The Wiley Blackwell companion to semantics*. Oxford: John Wiley & Sons. DOI: 10.1002/9781118788516.sem057.

Homer, Vincent. 2021. Domains of polarity items. *Natural Language Semantics* 38. 1–48. DOI: 10.1093/jos/ffaa006.

Horn, Laurence R. 1989. *A natural history of negation*. Chicago: Chicago University Press.

Horn, Laurence R. 2000. Pick a theory (not just *any* theory). In Laurence R. Horn & Yasuhiko Kato (eds.), *Negation and polarity: Syntactic and semantic perspectives*, 147–192. Oxford: Oxford University Press.

Howard, Edwin M. 2014. *Superlative degree clauses: Evidence from NPI licensing*. Massachusetts Institute of Technology. (Doctoral dissertation).

Israel, Michael. 1998. *Ever*: Polysemy and polarity sensitivity. *Linguistic Notes from La Jolla* 19. 29–45.

Jeong, Sunwoo & Floris Roelefsen. 2023. Focused NPIs in statements and questions. *Journal of Semantics* 40. 1–48. DOI: 10.1093/jos/ffac014.

Kadmon, Nirit & Fred Landman. 1993. *Any*. *Linguistics and Philosophy* 16. 353–422. DOI: 10.1007/bf00985272.

Karttunen, Lauri. 1971. Implicative verbs. *Language* 47. 340–358. DOI: 10.2307/412084.

Karttunen, Lauri. 1974. *Until*. In Michael W. La Galy, Robert A. Fox & Anthony Bruck (eds.), *CLS 10: Papers from the 10th Annual Meeting of the Chicago Linguistic Society*, 284–297. Chicago, Illinois: Chicago Linguistic Society.

Klima, Edward. 1964. Negation in English. In J. Fodor & J. Katz (eds.), *The structure of language*, 246–323. New Jersey: Prentice Hall.

Krifka, Manfred. 1995. The semantics and pragmatics of polarity items. *Linguistic Analysis* 25. 209–257.

Krifka, Manfred. 2010. *Before* and *After* without coercion: Comment on the paper by Cleo Condoravdi. *Natural Language and Linguistic Theory* 28. 911–929. DOI: 10.1007/s11049-010-9116-y.

Kroll, Margaret. 2019. Polarity reversals under sluicing. *Semantics and Pragmatics* 12. DOI: 10.3765/sp.12.18.

Kroll, Margaret & Deniz Rudin. 2018. Identity and interpretation: Syntactic and pragmatic constraints on the acceptability of sluicing. In Andrew Lamont & Katerina Tetzloff (eds.), *NELS 47: Proceedings of the Forty Seventh Annual Meeting of the North East Linguistic Society*, vol. 2, 177–190. Amherst, Massachusetts: University of Massachusetts, Graduate Student Linguistic Association.

Labov, William. 1972. Negative attraction and negative concord. *Language* 48. 773–818.

Ladusaw, William A. 1979. *Polarity sensitivity as inherent scope relations*. University of Texas at Austin. (Doctoral dissertation).

Ladusaw, William A. 1992. Expressing negation. In Chris Barker & David Dowty (eds.), *SALT 2, Proceedings of the 2nd Semantics and Linguistic Theory Conference*, 237–259. Published as Volume 40 of the Ohio State University Working Papers in Linguistics. Columbus, Ohio: Ohio State University.

Lahiri, Utpal. 1998. Focus and negative polarity in Hindi. *Natural Language Semantics* 6. 57–123. DOI: 10.1023/A:1008211808250.

Laka, Itziar. 1990. *Negation in syntax: On the nature of functional projections in syntax*. Massachusetts Institute of Technology. (Doctoral dissertation).

Laka, Itziar. 1993. Negative fronting in Romance: Movement to Sigma. In William J. Ashby, Marianne Mithun & Giorgio Perissinotto (eds.), *Linguistic perspectives on Romance languages: Selected papers from the xxi linguistic symposium on Romance languages, Santa Barbara, February 21–24, 1991*, 315–334. Amsterdam: John Benjamins Publishing Company.

Lee, Chungmin. 1995. A unified account of polarity phenomena. In *Pacific Asia Conference on Language, Information and Computation*.

Lee, Young-Suk & Laurence R. Horn. 1994. *Any* as indefinite + *even*. Manuscript, Yale University.

LeGrand, Jean. 1975. *Or and any: The semantics of two logical operators*. University of Chicago. (Doctoral dissertation).

Linebarger, Marcia. 1987. Negative polarity and grammatical representation. *Linguistics and Philosophy* 10. 325–387.

Maki, Hideki & Dónall P. Ó Baoill. 2011. *Essays on Irish syntax*. Bunkyo-ku, Tokyo: Kaitakusha Co. Ltd.

Maki, Hideki & Dónall P. Ó Baoill. 2017. Embedded topicalization in Irish. In *Essays on Irish syntax II*, 139–157. Bunkyo-ku, Tokyo: Kaitakusha Co. Ltd.

Mayr, Clemens. 2013. Downward monotonicity in questions. In Emmanuel Chemla, Vincent Homer & Grégoire Winterstein (eds.), *Proceedings of Sinn und Bedeutung 17*, 345–362. Paris.

McCloskey, James. 1985. The Modern Irish double relative and syntactic binding. *Ériu* 36. 45–84.

McCloskey, James. 1996. On the scope of verb raising in Irish. *Natural Language and Linguistic Theory* 14. 47–104.

McCloskey, James. 2001. The morphosyntax of WH-extraction in Irish. *Journal of Linguistics* 37. 67–100. DOI: 10.1017/S0022226701008775.

McCloskey, James. 2002. Resumption, successive cyclicity, and the locality of operations. In Samuel Epstein & Daniel Seeley (eds.), *Derivation and explanation*, 184–226. London: Blackwell Publishers.

McCloskey, James. 2013. Focus and the grammar of exceptives in Irish. Paper presented at the 19th International Congress of Linguists, Geneva, July 22nd–27th 2013. Handout available at https://gramadach.net/mccloskey/papers.html. (5 June, 2025).

McCloskey, James. 2017. Ellipsis, polarity, and the cartography of verb-initial orders in Irish. In Enoch Aboh, Eric Haeberli, Genoveva Puskás & Manuela Schönenberger (eds.), *Elements of comparative syntax: Theory and description* (Studies in Generative Grammar 127), 99–151. Boston: De Gruyter Mouton.

Merchant, Jason. 2004. Fragments and ellipsis. *Linguistics and Philosophy* 27. 661–738. DOI: 10.1007/s10988-005-7378-3.

Merchant, Jason. 2010. Three kinds of ellipsis. In François Recanati, Isidora Stojanovic & Neftalí Villanueva (eds.), *Context-dependence, perspective, and relativity*, 141–192. Berlin: Walter de Gruyter.

Mittwoch, Anita. 1977. Negative sentences with *until*. In Woodford A. Beach, Samuel E. Fox & Shulamith Philosoph (eds.), *CLS 13: Papers from the 13th Annual Meeting of the Chicago Linguistic Society*, 410–417. Chicago, Illinois: Chicago Linguistic Society.

Ó Siadhail, Mícheál. 1989. *Modern Irish: Grammatical structure and dialectal variation*. Cambridge: Cambridge University Press.

Ostrove, Jason. 2018. Stretching, spanning and linear adjacency in vocabulary insertion. *Natural Language and Linguistic Theory* 36. 1263–1289. DOI: 10.1007/s11049-018-9399-y.

Penka, Doris & Hedde Zeijlstra. 2010. Negation and polarity: An introduction. *Natural Language and Linguistic Theory* 28. 771–786. DOI: 10.1007/s11049-010-9114-0.

Preminger, Omer & Maria Polinsky. 2015. Agreement and semantic concord: A spurious unification. Manuscript, University of Maryland.

Ranero, Rodrigo. 2021. *Identity conditions on ellipsis*. University of Maryland, College Park. (Doctoral dissertation).

Roberts, Craige. 2012. Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5. 1–69. DOI: 10.3765/sp.5.6.

Robinson, Mary & Gary Thoms. 2021. Long distance negative concord in English. In Alessa Farinella & Angelica Hill (eds.), *NELS 51, Proceedings of the Fifty First Meeting of the North East Linguistics Society, Volume Two*, 181–190. Amherst, Massachusetts: University of Massachusetts, Graduate Student Linguistic Association.

Rudin, Deniz. 2019. Head-based syntactic identity in sluicing. *Linguistic Inquiry* 50. 253–283. DOI: 10.1162/ling_a_00308.

Stainton, Robert. 2006. *Words and thoughts: Subsentences, ellipsis, and the philosophy of language*. Oxford: Oxford University Press.

Toosarvandani, Maziar. 2013. Corrective *but* coordinates clauses not always but sometimes. *Natural Language and Linguistic Theory* 31. 827–863. DOI: 10.1007/s11049-013-9198-4.

Valmala, Vidal. 2007. The syntax of little things. In Yehuda N. Falk (ed.), *Proceedings of Israel Association for Theoretical Linguistics*, vol. 23. Tel Aviv University.

van der Wouden, Ton. 1994. *Negative contexts*. Rijksuniversiteit Groningen. (Doctoral dissertation).

van der Wouden, Ton. 1997. *Negative contexts. Collocation, negative polarity, and multiple negation*. London: Routledge.

von Fintel, Kai. 1999. NPI licensing, Strawson entailment, and context dependency. *Journal of Semantics* 16. 97–148. DOI: 10.1093/jos/16.2.97.

Weir, Andrew. 2014. *Fragments and clausal ellipsis*. University of Massachusetts Amherst. (Doctoral dissertation).

Weir, Andrew. 2015. Fragment answers and 'exceptional movement under ellipsis': A PF-movement account. In Thuy Bui & Deniz Özyıldız (eds.), *NELS 45, Proceedings of the Forthy Fifth Meeting of the North East Linguistics Society*, 175–188. Amherst, Massachusetts: University of Massachusetts, Graduate Student Linguistic Association.

Yoshida, Masaya. 2010. Antecedent-contained sluicing. *Linguistic Inquiry* 41. 348–356. DOI: 10.1162/ling.2010.41.2.348.
Zanuttini, Raffaella. 1997. *Negation and clausal structure*. Oxford: Oxford University Press.
Zeijlstra, Hedde. 2004. *Sentential negation and negative concord*. University of Amsterdam. (Doctoral dissertation).
Zeijlstra, Hedde. 2008. Negative concord is syntactic agreement. Manuscript, University of Amsterdam.
Zeijlstra, Hedde. 2012. There is only one way to Agree. *The Linguistic Review* 29. 491–539. DOI: 10.1515/tlr-2012-0017.
Zeijlstra, Hedde. 2016. Negation and negative dependencies. *Annual Review of Linguistics* 2. 233–254. DOI: 10.1146/annurev-linguistics-030514-125126.
Zwarts, Frans. 1996. Three types of polarity. In Fritz Hamm & Edward W. Hinrichs (eds.), *Plural quantification*, 177–238. Dordrecht: Kluwer Academic Publishers.

Chapter 2

When linearization triggers embedded V2: Evidence from Breton

Melanie Jouitteau
Centre National de la Recherche Scientifique (CNRS)

In this chapter, I present a subset of Breton embedded verb-second word orders. These data are intriguing because they implicate operations in embedded domains which appear to be optional. By contrast, we know that in matrix clauses the same operations are last resort, non-optional ones. I show that a proper ordering of operations at PF can ascribe optionality to a dialect parameter that allows for the exclusion of high functional heads from the start of a linearization domain. If and only if a functional head is excluded, would we observe last resort operations for T2. This ordering of operations is significant for our study of the modularity of grammar, because linearization of the syntactic structure can be shown to precede a set of morphological operations which operate in a linear domain.

1 Introduction

Unlike the restrictions against verb second (V2) in languages like German, Breton allows embedded verb second orders in some circumstances. The matrix word order in Breton is consistently verb-second: the tensed element typically follows another constituent. I will refer to the phenomena of ordering in Breton as Tense second (T2) rather than V2 because in compound tenses, the second position is defined relative to Finiteness and Tense. Consider the data in (1).[1] In the matrix clause of (1a), the subject *tout an dud* 'everybody' precedes a tensed lexical verb.

[1]In this chapter, the finite element subject to second position placement is highlighted in *italic*, while the element fulfilling the first position is in **bold**. Trace copies are indicated with ~~strikethrough~~.

Melanie Jouitteau. 2025. When linearization triggers embedded V2: Evidence from Breton. In Andrew Carnie, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond & Luis Irizarry (eds.), *Foundational approaches to Celtic linguistics*, 63–95. Berlin: Language Science Press. DOI: 10.5281/zenodo.15654859

This subject can be replaced by the object *keuz* 'regret', an adverb, or even a syntactic head. In the embedded domain of (1a), the tensed element is the second constituent following a complementizer, which lacks a phonological matrix. This is not surprising – it aligns with extensive evidence that the T2 requirement may be satisfied by a syntactic head, such as negation, or by a phonologically empty element such as an omitted topic. The alternate word order in (1b) is particularly intriguing. It involves an additional derivational step through stylistic fronting. The puzzle lies in the apparent optional nature of participle head movement in the embedded clause. Normally, a defining characteristic of stylistic fronting in Breton is its role as a last resort mechanism to ensure the second placement of the tensed element.

(1) a. **Tout an dud** *o* *d-eus* keuz **∅** *e* *n-eus* lennet
all the people FIN.3PL 3-AUX regret C FIN 3-AUX.PRES read.PTCP
Alain an dra-se.
Alain the thing-there
'Everyone regrets that Alain read this thing.'

b. **Tout an dud** *o* *d-eus* keuz ∅ **lennet** *e*
All the people FIN.3PL 3-AUX.PRES regret C read.PTCP FIN
n-eus ~~lennet~~ Alain an dra-se.
3-AUX.PRES ~~read.PTCP~~ Alain the thing-there
'Everyone regrets that Alain read this thing.'

The data from Breton were gathered through a series of 14 elicitation sessions conducted between 2015 and 2018, involving three traditional native speakers from distinct Breton dialects. These speakers, all over 70 years of age, were raised primarily speaking Breton with French as their early second language. They are referred to by their chosen names or acronyms. The first speaker, Huguette Gaudart from Skaer and Bannaleg, is proficient in both the East Kerne dialect and Standard Breton. The other two speakers, MLB from Plougerne and AM from Lesneven/Kerlouan, each represent a point of variation in the Leon dialect. They both have minimal influence from other dialects, including Standard Breton. At the time of the elicitation, the speakers were already familiar with the author from previous sessions and were accustomed to providing grammaticality judgments. The elicitation protocol involved translation tasks from French. The responses were transcribed in front of the speakers using the orthography of their preference. Following this, the speakers were asked to evaluate the grammaticality of various word permutations that altered the sentence they had originally provided. Comprehensive documentation of the raw elicitation data, along with

a detailed description of the elicitation processes, is available online at the Elicitation Center of the wikigrammar ARBRES (Jouitteau 2007-2024).

The structure of this chapter is as follows. First, I establish a theoretical framework for the analysis of T2 in Breton. This includes a detailed description of the architecture of the left periphery in Breton clauses, along with a summary of relevant research on the phenomena of second position placement within the matrix domain. Section 3 introduces the pivotal dataset. I present novel data from the Breton Kerne and Leon dialects, documenting a range of unexpected embedded V2 orders. A complementizer is always acceptable as the first element of the embedded (C-VSO), but sometimes in Leon we also find intervening heads (C-X-VSO). The paradigms they come from are last resort for T2 in matrix domains (expletive insertion, stylistic fronting, analytic tenses). I show the limitations on these rare data before turning to their analysis. Section 4 presents, and subsequently refutes, the hypothesis that embedded T2 orders can be fundamentally considered as root clauses with a relaxed syntactic integration. This section demonstrates that the embedded T2 domain exhibits weak island effects. Our arguments are supported by evidence from the C-condition, quantifier binding, and the deictic interpretation of temporal nouns. On these grounds, I reject the hypothesis of Wurmbrand (2012, 2014): embedded T2 domains are fully integrated into the syntactic structure. Section 5 examines matrix and embedded evidence for the linearization of high complementizers and clause external heads. Remarkably, even Breton clause external heads can be mobilized to pass the T2 filter.

2 Breton in the typology of V2 languages

The sentences in example (2) illustrate a classic paradigm for second-position phenomena. A constituent occupies the initial position of the clause, followed by the inflected element, with the remaining arguments succeeding these. These sentences are semantically equivalent, differing only in their discourse effects. Breton hence qualifies as a V2 or T2 language (Stephens 1982), a class of languages that consistently show the same paradigms (for a detailed typology of those languages, see Jouitteau 2010).

(2) a. **Warc'hoazh** *e* *preno* Yann ul levr d'am c'hoar ~~Warc'hoazh~~.
tomorrow FIN buy.FUT.3S Yann a book to my sister ~~tomorrow~~
'Yann will buy a book for my sister TOMORROW.' ADV

b. **Ul levr** *a* *breno* Yann ~~ul levr~~ d'am c'hoar warc'hoazh.
a book FIN buy.FUT.3S Yann ~~a book~~ to my sister tomorrow
'Yann will buy A BOOK for my sister tomorrow.' O

c. **Yann** *a* *breno* ~~Yann~~ ul levr d'am c'hoar warc'hoazh.
Yann FIN buy.FUT.3S ~~Yann~~ a book to my sister tomorrow
'YANN will buy a book for my sister tomorrow.' S

d. Warc'hoazh, **Yann** *a* *breno* ~~Yann~~ ul levr d'am c'hoar
tomorrow Yann FIN buy.FUT.3S ~~Yann~~ a book to my sister
~~Warc'hoazh~~.
~~tomorrow~~
'Yesterday, YANN will buy a book for my sister tomorrow.' ADV, S

Within the typology of T2 languages, the Breton data reveal certain specific features such as a wide left periphery (2d), and some Celtic-specific features such as argument ordering (in the middle field, subjects precede objects, VSO), and a particle, either *a* or *e*, that consistently precedes the tensed element. It is glossed here as FIN, the head of the Finite projection. Unlike Romance or Germanic languages then, the lowest head in the articulated CP periphery is morphologically realized in Breton. This facilitates the structural placement of the finite verb inside the syntactic structure, in both matrix and embedded sentences.

2.1 The structure of the left periphery

The FIN particle invariably precedes only the tensed element, and is absent before infinitives, participles, or in small clauses. It manifests solely as *a* /a/ or *e* /e/. However, higher complementizers, such as negation, may fuse with it. In Standard Breton, as illustrated above, these particles alternate based on the ±nominal character of the preceding element (Rezac 2004). For example, in (2b,2c,2d), the particle *a* follows a nominal element, while in (2a), the particle *e* follows a non-nominal element. These particles are typically unstressed and often omitted in normal speech tempo, though speakers are aware of the unpronounced particle. Each particle, whether pronounced or not, induces a specific mutation in the subsequent word. For example, the particle *a* in (2b,2c,2d) consistently triggers lenition, changing the initial consonant of the inflected verb *prenañ* 'to buy' to a /b-/. Conversely, the particle *e* in (2a) triggers a set of mixed mutations, leaving the initial /p-/ in this case unaffected.

Another contrast with Germanic T2 languages is that the Breton verb consistently raises in all tensed sentences, resulting in embedded VSO in the simplest

cases. Importantly, the phonological realization of a C head above the verb does not influence verb raising. Verb raising occurs not only in T2 constructions but also in the (comparatively rare) Tense initial sentences in Breton, as well as in embedded clauses. The analysis of the trigger for T2 requirement is delineated in Jouitteau (2020), under the term of Left Edge Filling Trigger. Here, I present only the properties that are relevant for the study of embedded domains. The filter applying at linearization is in (3).

(3) A FIN head can be linearized if and only if it is not initial.

The domain to be linearized is maximally structured as illustrated in Figure 1 (page 68). In matrix sentences, only a coordination marker can be merged above the C line in Figure 1. The region between the C and B lines represents the higher projections of the CP area. This area can be empty or heavily populated. It is recursive. It may include a hanging topic, often followed by a prosodic break. It can also encompass adverbs and clauses that serve as scene-setting material as in (2d). The T2 requirement is entirely blind to this high left periphery located between the C and B lines. Empirically, hanging topics are never the only element preceding the tensed one. Scene-setting adverbs can, but only when they are focalized, and thus positioned below the B line. Placement below the B line obtains a discourse difference and restricts scene-setting adverbs to a focus reading. The expansive space between the B and C lines in Breton permits variations up to T6 orders. This makes the Breton T2 distinctly different from the strict T2 order found in Norwegian. Consequently, Breton is categorized as an "at least T2" type, similar to the patterns observed in Old Romance, Rhaeto-Romance, Karitiana (Arikém), and the Germanic languages Mòcheno or Cimbrian. This categorization is supported by the work of Bidese et al. (2014), as well as Jouitteau (2010; 2007-2024), who discuss the concept of Breton T3 orders and provide relevant references. In embedded domains, this area is expected to be reduced. Every element situated below this, between the B line and the A line, engages with the T2 requirement and can be the only pronounced element before a tensed head, as illustrated in the previous examples. Notably, in Breton, the syntactic heads located between the B line and the A line can also function as the sole initial element of a sentence. This property is crucial for the inquiry of embedded domains.

2.2 Breton T2 is linear

Breton instantiates *linear verb second*, as described by Borsley & Kathol (2000), Jouitteau (2020) and references therein. This terminology underlines the fact that

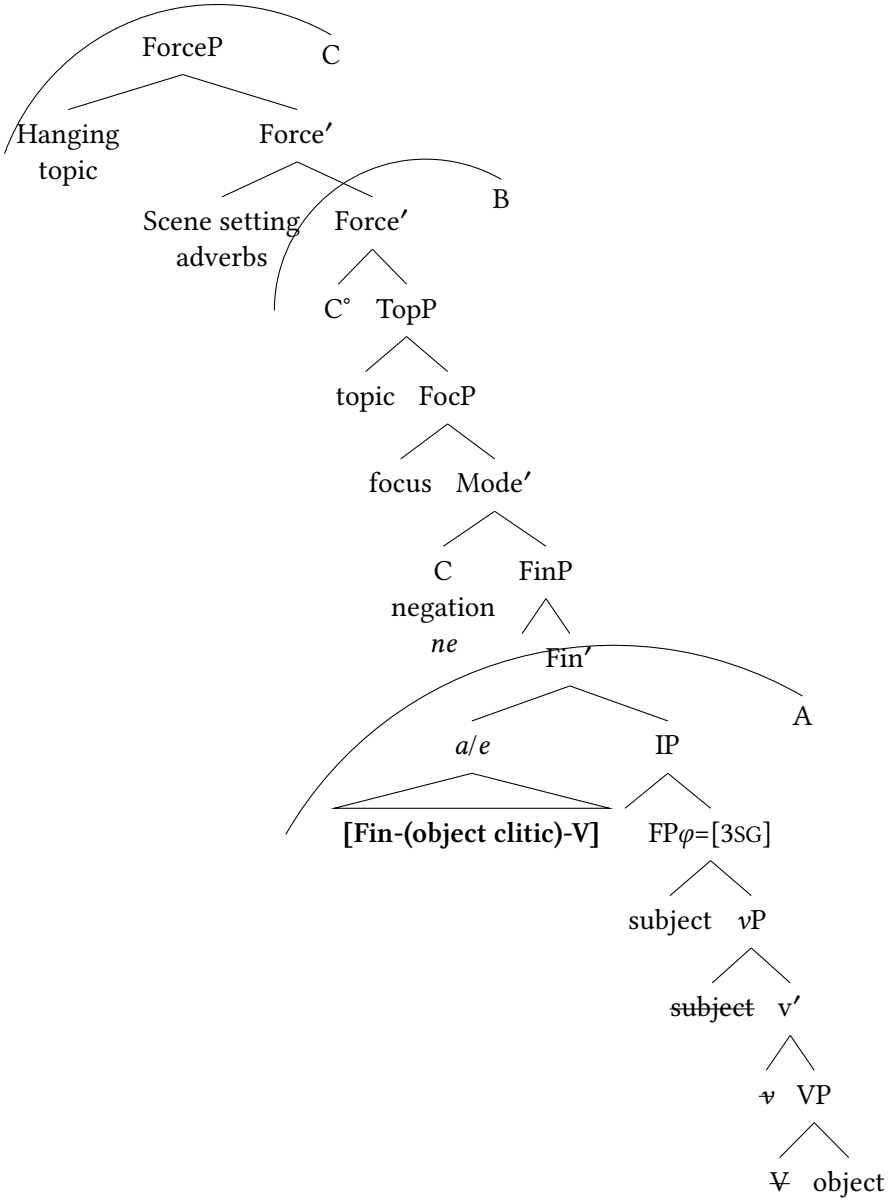


Figure 1: Structure of the Breton left periphery

the T2 requirement is blind to the [±head] distinction. Initial functional heads can be Q particles (4), preverbal negation (5), and even some C particles which are integrated into the verbal morphology (6). The latter case is predominantly observed with the verb *emañ*, which raises beyond the Fin head, reaching up to the C head of the ForceP projection. Additionally, a rare instance of verb-first word order is given in (7). In that case, an epenthetic *h-* is present at the beginning of the verbal root. This is likely the morphological residue of an ancient epenthetic consonant, induced by a complementizer that ends in a vowel, as illustrated in the phrase (*Ec'h an da …* 'I'm going to …').[2]

(4) **Hag** *ez* *eo* gwir an dra-se?
Q FIN AUX.3S true the thing-here
'Is this true?'

(5) Yann ha Lisa **ne** *brenint* ket al levr d'am c'hoar warc'hoazh.
Yann and Lisa NEG buy.FUT.3PL NEG the book to my sister tomorrow
'Yann and Lisa will not buy the book for my sister tomorrow.'

(6) **E** *mañ* Yann ha Lisa o prenañ ul levr d'am c'hoar.
C.be.3S Yann and Lisa at buy.INF a book to my sister
'Yann and Lisa are buying a book for my sister.'

(7) **'H** *an* da brenañ ul levr d'am c'hoar.
C.go.1S to buy.INF a book to my sister
'I will buy a book for my sister.'

Thus far, we have seen only examples where the element before Tense was either introduced from the numeration (e.g., a Q head, a Neg head, or a Force head when it is morphologically integrated into a verb), or by an obligatory movement due to the information structure (e.g., preverbal movement of a *wh-* phrase or another focalized element, or generation of a topic). The most striking property of the Breton T2 requirement is observed when none of those elements appear in the numeration. This context reveals a wide array of last resort strategies obtaining T2.

2.3 Markers of last resort operations

Positive, affirmative, matrix wide focus sentences (8–11) provide the proper context for observing the last resort operations. The sentence in example (8) demonstrates expletive insertion, where the expletive *Bez'* manifests as a morphological

[2]See also the matrix verb of (16) below.

abbreviation of the verb 'to be'. In (8), none of the verb's arguments are omitted, a sign that we are observing a transitive expletive. Example (9) shows the productive paradigm of stylistic fronting (also referred to in the literature on Slavic languages as long head movement) characterized by a fronted participle head crossing over a finite auxiliary in the derivation (Borsley et al. 1996 and references therein). The structure in example (10) represents an analytical tense using dummy 'do' insertion, while the structure in example (11) is essentially analogous to (10) but featuring verb-doubling.[3]

(8) **Bez'** *en* *deus* prenet Yann ul levr d'am c'hoar.
EXPL FIN.3SM aux.3S buy.PTCP Yann a book to my sister
'Yann has bought a book for my sister.'

(9) **Prenet** *en* *deus* ~~prenet~~ Yann ul levr d'am c'hoar.
buy.PTCP FIN.3SM AUX.3S ~~buy.PTCP~~ Yann a book to my sister
'Yann has bought a book for my brother.'

(10) **Prenañ** *a* *ra* Yann ul levr d'am c'hoar.
buy.INF FIN do.PRES.3S Yann a book to my sister
'Yann buys a book for my brother.'

(11) **Gouzout** *a* *ouzon* ar wirionez.
know.INF FIN know.PRES.1S the truth
'I know the truth.'

All the structures in (8) to (11) appear in pragmatic contexts lacking focus on an individual constituent.[4] Each serves as a valid response to questions of the *What happened?* variety. However, in terms of information structure, the sentences under consideration are not strictly equivalent. The structures observed in examples (9) and (10) are obligatorily backgrounded. In contrast, the structures in examples (8) and (11) typically exhibit a *verum focus*. This focus emphasizes the truthfulness of the entire sentence. The scope of this focus encompasses the entire sentence.

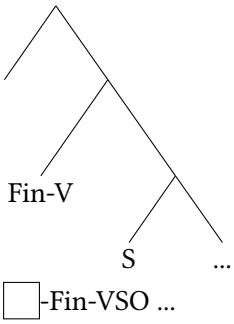
The sentences presented in examples (8) to (11) are all derived from mechanisms employed as a last resort to satisfy the T2 requirement. This requirement

[3] Jouitteau (2012) derives (10) and (11) by excorporation of the lexical verb out of the Fin complex, followed by erasure (10) or pronunciation (11) of the lowest copy. The lexical verb in these analytical tenses in never found lower in the structure, after the tensed element, which is support for the last resort excorporation analysis.

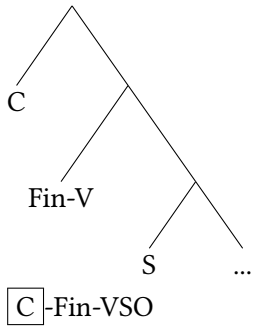
[4] More examples are provided in Jouitteau (2007-2024) on the pages on 'Bez' and 'verb-doubling'.

is characterized as a prohibition against placing T or Fin at the beginning of a linear sentence, as represented in (12a). The hypothesis that the mechanism filling the empty position in (12a) is a last resort strategy whose various results are observed in (8) to (11), leads to several accurate predictions. The structures observed in (8) to (11) are incompatible with focus of any argument, with topic, or with negation, banning of course *wh-* questions. Furthermore, each structure in (8) to (11) is mutually exclusive with the others.

(12) a. matrix clause



b. embedded clause



Under the assumption that embedded domains begin with a complementizer, this complementizer should uniformly fulfill the Tense-second (T2) linear requirement, as shown in the tree in (12b). Empirically speaking, documenting embedded domains with any structures akin to those in examples (8) through (11) would imply that the complementizer, for some reason, failed to linearize in the expected position, which in turn triggered last resort T2 operations. Understanding the underlying reasons for this phenomenon could markedly advance our comprehension of the embedding process in grammatical theory. The subsequent section will summarize prior findings on the grammar modules at play.

2.4 Post-syntactic morphology and linearization

We have evidence suggesting that the operation responsible for positioning the initial element in its place in examples (8) to (11) occurs in a morphological module whose input is the output of syntactic operations. The last resort effects are invisible for interpretation. They violate syntactic rules. Specifically, the initial element in example (9) contravenes the head-movement constraint (Travis 1984) by allowing a head to cross over another head in the derivation process. In example (8), merge of the transitive expletive has no impact on interpretation, not even indirectly such as a restriction on the definiteness of an argument. In (10),

pronunciation of the infinitive verb before Tense makes no difference with a synthetic verb on the semantic interface. Auxiliary insertion is equally imperceptible to semantic interpretation, and independently suggests a morphological module. Lastly, verb-doubling, as observed in example (11), is confined to a specific set of verbs that do not constitute a syntactic class, and show morphological idiosyncrasies, as argued by Jouitteau (2020).[5] Rivero (1999, 2000) first proposed that Breton Tense-second results from a filter operating after syntax. She identified this module as the phonological form (PF). This hypothesis is on the right track but is empirically incorrect: the last resort operations for T2 are followed by morphological operations. The mechanism responsible for the initial element is located in a grammatical module capable of processing categorical distinctions. The preverbal particle, glossed as FIN in our examples, is responsive to the [± nominal] category of the initial element.[6]

In prior research (Jouitteau 2011, 2012, 2020), I have suggested that the T2 requirement amounts to morphological exponence operating at the clause level. This requirement operates after syntax, in a module where non-syntactic operations can be dealt with, such as excorporation, inversion, idiosyncrasy, auxiliary insertion, copy deletion and non-deletion. These operations are extremely local (inversion, excorporation), and linear in nature.

The derivation of Breton clauses is recapitulated in (13). Breton is typologically consistent with the other Celtic languages in its syntax: at the level of syntax, Breton is VSO like any other Celtic language. A language-specific morphological requirement operating at the level of the clause, historically gained by contact with Romance T2, imposes that Fin is not initial.

(13) a. Breton has an obligatory morphological exponence operating at the level of clauses.
b. It applies to both matrix and embedded clauses.
c. This filter operates after syntax and before PF, in a module that can manipulate [±nominal] categories, proceed to linear inversion of elements, apply excorporation and handle morphological support like external merge of dummy auxiliaries or expletives.

[5]Verb-doubling is also idiosyncratic. Speakers differ in the verbs they double. Not all speakers are doublers. The Leon speaker MLB for example never does.

[6]This phenomenon is not uniformly evident across all dialects. Only in the dialects of Leon, Standard, and Gwened do the Fin heads consistently alternate based on the [± nominal] category of the preceding constituent. In this article, embedded T2 orders are from Leon and show robust alternations depending on categorical distinctions.

The generalization in (13c) leads to the speculation that the last resort operations for T2 happen during the linearization process of the FIN phrase. The invisibility of hanging topics and scene-setting adverbs between the B and C lines in Figure 1 is tentatively correlated with the sensitivity of this linearization process to prosody. If the linearization process is prosody sensitive, then we could correctly obtain the desired prediction that the obligatory exponence is blind to the domains that are prosodically isolated. I leave this for further research. We now are equipped to step into the embedded domains.

3 The trouble with linear T2 embeddings

Embedded domains are initiated by a C head. Our observations indicate that Breton functions as a linear T2 language wherein functional heads can fulfill T2 requirements. Consequently, it is predicted that the linearization of wide focus embedded sentences should consistently follow a C-Fin-VSO pattern.[7] Consequently, employing last resort T2 strategies should be uniformly ungrammatical across these embedded domains.

This prediction generally holds true, and definitely does in standard Breton. Our Kerne speaker Huguette Gaudart strongly opposes any form of embedded stylistic fronting, as exemplified in example (14) – here with a phonologically null Fin head. Her judgments apply both to her native Skaer dialect of Kerne and to her proficient standard dialect of the language. This generalization has been found consistently valid for her across various types of complementizers. Her C-FIN-VSO orders never accommodate optional alternations with structures such as presented in examples (8) to (11).

(14) Kerne (Skaer/Banaleg) (Gaudart 04/2016b)

a. N' uion ket **ma** ' *n-eus* lennet al levr.
C.NEG know.PRES.1S NEG if FIN 3-AUX read.PTCP the book
'I don't know if she has read the book.'

b. * N' uion ket ma **lennet** ' *n-eus* ~~lennet~~ al
C.NEG know.PRES.1S NEG if read.PTCP FIN 3.AUX ~~read.PTCP~~ the
levr.
book
'I don't know if she has read the book.'

[7] Some embedded domains are known to show an exceptionally wide left periphery, like those headed by *rak* 'because', and those which tolerate focus or even hanging topics hosted in their wide left periphery. Their word order is basically that of root clauses. These complementizers never lead to C-Fin-VSO orders. We will not consider these structures here.

Surprisingly, the speakers of the Leon dialect exhibit sporadic exceptions and allow for structures equivalent to (14b). In the remainder of this section, we will first examine these exceptions. We then address the limitations of this evidence.

3.1 Evidence for embedded expletive strategies

In example (15a), the word order resembles that in (14), where the *ma* complementizer leads a conventional linear T2 order. In contrast, example (15b) exhibits stylistic fronting, also grammatical for the speaker. Some caution is in order in this specific example because my conclusion is dependent on the generalization that the speaker has Aux-V-S orders (*Ma ʻn-eo renket an amezog…*) before inversion of the auxiliary and the verb. The grammaticality of this specific structure however could not be empirically verified during the research protocol: the speaker gave Aux-S-V in (15a). The analysis that (15b) involves stylistic fronting proper and not remnant is predicated on the assumption that the past participle has undergone derivation at an intermediary stage, positioned immediately after the Tense head. This assumption however appears here plausible, given that this speaker consistently permits participles to precede a lexical subject in the middle field in other instances.

(15) Leon (Plougerne) (MLB 05/2018)

a. **Ma** ʻ *n-eo* an amezog renket e wele ne hellan ket
if FIN 3S-AUX.FUT the neighbor tidy.PTCP his bed NEG can.1S NEG
lavaret dit!
tell.INF you
'Whether the neighbor will have made his bed, I can't tell you!'

b. Ma **renket** ʻ *n-eo* ~~renket~~ an amezog ~~renket~~ e
if tidy.PTCP FIN 3S-AUX.FUT ~~tidy.PTCP~~ the neighbor ~~tidy.PTCP~~ his
wele ne hellan ket lavaret dit!
bed NEG can.1S NEG tell.INF you
'Whether the neighbor will have made his bed, I cannot tell you!'

The sentences in (16) replicate the findings with the tense complementizer *pa* 'when'.

(16) Leon (Plougerne) (MLB 05/2018)

a. **Pa** *m-eus* kleet ar janson-se evit ar wech kentañ,
When 1S-AUX hear.PTCP the song-there for the time first
eved-on dougeres.
C.AUX.PAST-1S pregnant
'When I first heard this song, I was pregnant.'

b. Pa **kleet** *m-eus* ~~kleet~~ ar janson-se evit ar wech
when hear.PTCP 1S-AUX ~~hear.PTCP~~ the song-there for the time
kentañ, eved-on dougeres.
first C.aux.PAST-1S pregnant
'When I first heard this song, I was pregnant.'

The sentences in (17) replicate the findings with another Leon speaker and another last resort strategy: an analytic tense (9).

(17) Leon (Lesneven/Kerlouan) (AM 05/2016)

a. Bevañ a r-eomp un amzer **hag** *e hoar* ar vugale
live.INF FIN do-PRES.1PL a time that FIN know.3S the child.PL
diouzh an ordinatourien muioc'h eget o zud.
of the computers more than their parents
'We live a time where children know more about computers than their parents.'

b. Bevañ a r-eomp un amzer ha **gouzout** *a r-a* ar
live.INF FIN do-PRES.1PL a time that know.INF FIN do-3S the
vugale diouzh an ordinatourien muioc'h eget o zud.
child.PL of the computers more than their parents
'We live a time where children know more about computers than their parents.'

The following examples replicate the findings with a declarative C head, likely the head of ForceP. This element is phonologically null (∅). This silent complementizer typically selects clausal complements in most Breton dialects, creating an apparent radical Tense-first order on the surface. The sentence in (18a) demonstrates that this C head alone can fulfill the T2 requirement. Embedded stylistic fronting is also a possibility, as shown in (18b). Within the clausal complement of a noun, as seen in (19), the C can either precede the verb directly or allow the insertion of an expletive, as in (19a, where the *bed* morpheme is a dialectal variation of the transitive expletive *bez'* (previously mentioned in (8)). An analytic

tense form, similar to what was illustrated in (10), represents yet another option, as seen in (19b) and in (20).

(18) a. Amelia e n-eus zouchet **∅** *en* *n-eus* kemeret an
Amelia FIN 3S-AUX think.PTCP C FIN.3S 3S-AUX take.PTCP the
alc'hwez fall.
key wrong
'Amelia thought she had taken the wrong key.'

b. Amelia e n-eus zouchet ∅ **kemeret** *en* *n-eus* ~~kemeret~~
Amelia FIN 3S-AUX think.PTCP C take.PTCP FIN.3S 3S-AUX ~~take.PTCP~~
an alc'hwez fall.
the key wrong
'Amelia imagined/thought she had taken the wrong key.'

(19) a. Ar brud vat e-n d-oa ∅ (**bed**) *e* *houie*
the reputation good FIN-3S 3-AUX.PAST C expletive FIN know.PAST.3S
dresañ pep tra.
fix.INF every thing
'He had a good reputation for knowing how to fix anything.'

b. Ar brud vat e-n d-oa ∅ **gouzout** *a* *r-ae*
the reputation good FIN-3S 3-AUX.PAST C know.INF FIN do-PAST.3S
dresañ pep tra.
fix.INF every thing
'He had a good reputation for knowing how to fix anything.'

(20) Leon (Plougerne) (MLB 05/2018)
Simone a jouch ∅ **gouzout** *a* *r-aio* ar sodenn-se troc'hañ
Simone Fin think C know.INF FIN do-FUT.3S the idiot-here cut.INF
uheloc'h ar harzenn.
higher the hedge
'Simone thinks that this idiot will know to cut the hedge higher.'

The other Leon speaker AM also permits the optional inclusion of embedded stylistic fronting, as demonstrated in example (21). This is illustrated with a past participle in (21b). Linear inversion can target any element immediately following the tensed element, as shown with an extra aspectual marker in (21c).

(21) Leon (Lesneven/Kerlouan) (AM 05/2016)

a. Mimi a joñch **∅** *e trouc'ho* ar sodenn-se uheloc'h ar
Mimi FIN think.PRES.SG C FIN cut.FUT.3S the idiot-here higher the
harzenn e-giz ar bloavezh paseet.
hedge like the year pass.PTCP
'Mimi thinks that this idiot will cut the hedge higher than last year.'

b. Mimi a joñch ∅ **troc'het** *n-eus* ~~troc'het~~ ar sodenn-se
Mimi FIN think.PRES.SG C cut.PTCP 3S-AUX ~~cut~~ the idiot-here
uheloc'h ar harzenn e-giz ar bloavezh paseet.
higher the hedge like the year pass.PTCP
'Mimi thinks that this idiot has cut the hedge higher than last year.'

c. Mimi a joñch ∅ **bet** *n-eus* ~~bet~~ troc'het ar
Mimi FIN think.PRES.SG C be.PTCP 3S-AUX ~~be.PTCP~~ cut.PTCP the
sodenn-se uheloc'h ar harzenn e-giz ar bloavezh paseet.
idiot-here higher the hedge like the year pass.PTCP
'Mimi thinks that this idiot has cut the hedge higher than last year.'

This study is not the first to report stylistic fronting as permissible within embedded domains in Breton. Rivero (1999) presented the critical minimal pair in (22), based on the grammatical judgment of Janig Stephens, a native speaker of the Treger dialect. The *-eñ* morpheme is homophone with the third-person singular masculine independent pronoun and forms a unique compound *hag-eñ*, which is exclusive to the Treger dialect. It occurs solely within embedded domains. The prevailing hypothesis, put forth in Lambert (1998: 318) and Rivero (1999: 81–82), posits that *-eñ* in example (22a) functions as an expletive. Notably, the embedded expletive *-eñ* and the stylistic fronting operation illustrated in example (22b) are observed to be in complementary distribution.

(22) Treger (Rivero 1999: 81–82)

a. N' ouz-on ket hag-**eñ** *e-n d-eus* lennet al
NEG know.PRES-1S NEG if-EXPL FIN-3SG 3-AUX.PRES read.PTCP the
levr.
book
'I don't know if he has read the book.'

b. N' ouz-on ket ha **lennet** *e-n d-eus* ~~lennet~~ al
NEG know.PRES-1S NEG if read.PTCP FIN-3S 3-AUX.PRES ~~read.PTCP~~ the

levr.
book
'I don't know if he has read the book.'

The empirical generalization that in the northern dialects of Leon and Treger, the initial presence of a C head in embedded domains does not ban post-syntactic word order re-arrangements captures these examples. Operations that we know to be last resort for T2 in matrix sentences here seem optional. Such options are limited, however.

3.2 Limits to embedded last resort T2

There are significant limitations to embedded T2. First, and this is very important, it is absent from corpora.[8] Following the descriptive literature so far, it does not exist. Prescriptive or pedagogical texts in standard Breton would unambiguously mark it as ungrammatical. Embedded last resort T2 orders are obtained only in an elicitation context. However, its grammaticality among the speakers interviewed is strong. Prior to participating in this elicitation, these individuals underwent testing to ascertain their ability to evaluate grammatical structures accurately. I verified that their judgments remain consistent over time. When queried about the potential distinction between the two options, the speakers reported no perceivable difference. Nevertheless, when asked to express a preference between the two, there was a clear and consistent inclination towards the regular C-Fin-VSO word orders.

Subtle variations in information structure significantly affect the grammaticality of embedded expletive strategies. Consider example (23), where MLB does not permit the stylistic fronting of the participle (23b). This phenomenon unlikely results from an intervention effect related to inversion, as this speaker typically allows for the positioning of past participles before lexical subjects. In contrast, subject fronting is deemed grammatical in (23c). The analysis in this case must consider the possibility that the speaker is invoking a context that induces a contrastive focus on the entity 'Julie'. Replication of this inquiry necessitates a rigorously controlled protocol attentive to information structure nuances and

[8] Breton corpora have reasonable size and quality, but the lack of a richly annotated corpus and parser impacts its searchability. The wikigrammar ARBRES (Jouitteau 2007-2024) describes Breton syntactic variation with hand-picked examples coming from about 500 published references. No embedded verb-second has been reported in syntactic environment where it is last resort, despite a long-term focus on the matter (Jouitteau 2007-2024: 'Embedded V2').

well-informed about the specific potential prosodies in both matrix and embedded sentences. For the moment, we are here facing a gap in the availability of prosodic models for Breton that would enable reliable predictions.

(23) Leon (Plougerne) (MLB (05/2018)

a. Me chouj d'ar poent-se e houie **∅** *e*
I think.PRES.1S at the time-there FIN know.PAST.3S C FIN
n-oa Julie kemeret va oto.
3-AUX.PAST Julie take.PTCP my car

b. * Me chouj d'ar poent-se e houie ∅ **kemeret**
I think.PRES.1S at the time-there FIN know.PAST.3S C take.PTCP
e n-oa Julie ~~kemeret~~ va oto.
FIN 3-AUX.PAST Julie ~~take.PTCP~~ my car

c. Me chouj d'ar poent-se e houie ∅ **Julie** *e*
I think.PRES.1S at the time-there FIN know.PAST.3S C Julie FIN
n-oa ~~Julie~~ kemeret va oto.
3-AUX.PAST ~~Julie~~ take.PTCP my car
'I think from that moment he knew that Julie had taken my car.'

At this stage, it is important to observe that alterations in prosody directly influence the interpretability of embedded clauses as independent root clauses. The introduction of prosodic breaks results in variations in grammaticality judgments. For instance, the speaker AM in (24a) deemed the embedded stylistic fronting as ungrammatical. However, the inclusion of a prosodic break, as shown in example (24b), rendered it grammatically acceptable. It should be noted, however, that this rectification through prosodic intervention is not universally applicable (25). The specific conditions where a prosodic break will make a sentence grammatical or not should be seriously investigated, with an informed modeling of Breton prosody.

(24) Leon (Lesneven/Kerlouan) (AM 05/2016)

a. * Me a gred ∅ **plantet** ' *n-eus* Simone patatez er
I FIN thinks C planted FIN 3.AUX.PRES Simone potato.PL in.the
bloavezh-mañ egist he amezog.
year-here like her neighbor
'I think that Simone has planted potatoes like her neighbor this year.'

b. Me a gred ∅ //**plantet** ‘ *n-eus* Simone patatez er
I FIN thinks C planted FIN 3.AUX.PRES Simone potato.PL in.the
bloavezh-mañ egist he amezog.
year-here like her neighbor
‘I think that Simone has planted potatoes like her neighbor this year.’

(25) Leon (Lesneven/Kerlouan) (AM 05/2016)
*Mimi a joñch ∅ //**gouzout** *a raio* ar sodenn-se troc’hañ
Mimi FIN thinks C know.INF FIN do.FUT.3S the idiot-here cut.INF
berroc’h ar harz e-giz ar bloavezh paseet.
shorter the hedge like the year passed
‘Mimi thinks that this idiot will know to cut the hedge shorter than last year.’

With this in mind, we will now check that in the gathered evidence for embedded T2 so far, the clauses were fully integrated into the syntactic structure, and not merely pronounced after another root clause. Our conclusion will be that embedded T2 sentences under investigation here are indeed syntactically embedded.

4 The syntactic integration of embedded T2

Wurmbrand’s (2012; 2014) work proposes an interface mechanism that accounts for variations in the integration of embedded T2 sentences. According to her model, if two clauses merge in syntax, they achieve full integration. Alternatively, they may merge in a module termed “Transfer” before being processed by Phonological Form (PF) and Logical Form (LF). The level of integration will next depend on Spell-out considerations. When realized as separate Spell-out domains, the two domains maintain complete independence. Conversely, if they are realized within a single Spell-out domain, their integration is partial. In such a case, phenomena such as quantifier binding are observable, yet the two structures retain syntactic independence. This is summarized in (26).

(26)
- **Fully integrated clauses**: Merge in syntax
- **Partially integrated clauses**: Merge at Transfer, single Spell-out domain
- **Non-integrated clauses**: Merge at Transfer, separate Spell-out domains

Wurmbrand (2014) relies on the root generalization, by which a TP or CP is a root clause if and only if it has interpretable tense in the top projection. This is meant to explain why in German the clauses that do not instantiate verb raising will never count as roots: German verb-final sentences will always be fully integrated embedded clauses. A stipulation bans Tense marked clauses Merging with a higher verb in syntax, and this leads to Merge outside of syntax, at *Transfer* before spell-out.

It remains unclear why in this scenario syntactic root clauses persistently start with a complementizer. However, the prediction is clear. In clauses that are tense-second on the surface but are only partially integrated into syntax or not integrated at all, we should document no extraction, no movement, and only root phenomena. I now show that Breton disproves those predictions. I propose that, in Breton and in probably all linear T2 languages, the word orders that surface as last resort T2 clauses are syntactically integrated. At the syntactic stage of the derivation they might still not be T2, and this order will be linearly enforced later in the linearization process.

4.1 Weak island effects

The control sentence in (27a) shows grammaticality of embedded stylistic fronting. The speaker showed some hesitation, and finally accepted the sentence. Extraction out of the clausal complements is grammatical for the adjunct *betek peleah* in (27b). Stylistic fronting is not compatible with this extraction in (27c).

(27) Leon (Plougerne) (MLB 05/2018)

a. ? Jack e n-eus chouch ∅ **baleet** *e n-eus* gad
Jack FIN 3S-AUX.PRES think.PTCP C walk.PTCP FIN 3S-AUX.PRES with
e jochonoù beteg an daol hep kouezhañ.
his slipper.PL to the table without fall.INF
'Jack remembers that with the slippers on, he walked to the table without falling.'

b. **Betek peleah** *e n-eus* Jack kredet **∅** [*e*
until where FIN 3S-AUX.PRES Jack believe.PTCP C FIN
n-eche baleet hep kouezhañ ~~betek peleah~~]?
3S-AUX.COND.PAST walk.PTCP without fall.INF ~~until where~~
'Until where did Jack think he would walk without falling?'

c. * **Betek peleah** *e n-eus* Jack kredet ∅ [**baleet** *e*
until where FIN 3S-AUX.PRES Jack believe.PTCP C walk.PTCP FIN
n-eche ~~baleet~~ hep kouezhañ ~~betek peleah~~]?
3S-AUX.COND.PAST ~~walk.PTCP~~ without fall.INF ~~until where~~
'Until where did Jack think he would walk without falling?'

Object extraction does not show the same restriction. Extraction out of the clausal complements is grammatical for the object *pese parti* in (28a). Stylistic fronting is compatible with this extraction in (28b). The grammaticality of (28b) shows this embedded T2 domain is syntactically integrated. Under the assumption that T2 was realized here by inversion of a postverbal subject into the pre-Tense position (stylistic fronting), (28b) shows that the embedded domains where we observe last resort operations for T2 are syntactically integrated. (29) replicates (28b) with a context that lowers the plausibility of the embedded subject being fronted by focus. The contrast between object extraction (28b, 29) and (27c) suggests that embedded T2 structures are weak islands.

(28) a. Pese parti ar journal a lavar ∅ [**e n-eus** an
what party the newspaper FIN say.PRES.3S C FIN 3S-AUX.PRES the
amezog soutenet ~~pese parti~~]?
neighbor support.PTCP ~~what party~~
'What party did the journal say that the neighbor has supported?'

b. Pese parti ar journal a lavar ∅ [**an amezog e**
what party the newspaper FIN say.PRES.3S C the neighbor FIN
n-eus ~~an amezog~~ soutenet ~~pese parti~~]?
3S-AUX.PRES ~~the neighbor~~ support.PTCP ~~what party~~
'What party did the journal say that the neighbor has supported?'

(29) Leon (Plougerne) (MLB (05/2018)
Petra an archerien e n-eus soñjet ∅ [**ar seurezed** *a*
what the policeman.PL FIN 3S-AUX.PRES think.PTCP C the sister.PL FIN
houie ~~ar seurezed~~]?
know.PAST.3S ~~the sister.PL~~
'What did the policemen think that the sisters knew (all along)?'

Similar data are available in Germanic languages, but their analysis is less straightforward because of parentheticals. It has been proposed for German that argument extraction from an embedded T2 clause results from the use of a parenthetical in a monoclausal *wh*-sentence (Reis 1997; Reis 2002; see Kiziak 2010

for a summary of this debate for Germanic languages). According to this perspective, the embedding in the example (30) should be analyzed more accurately as *Welchen film* [hat sie gesagt] *haben die Kinder gesehen.*

(30) Welchen film [hat sie gesagt [haben die Kinder gesehen]]?
what film AUX.3S she say.PTCP have the child.PL see.PTCP
'What film did she say that the children have seen?'

This counterargument does not apply to Breton. Breton lexical subjects in parentheticals are uniformly postverbal, as opposed to the lexical subjects illustrated above in (28b) or (29).

4.2 Freezing of clausal complements

Another syntactic piece of evidence comes from movement of clausal complements. In most dialects, including the standard, fronting of a clausal complement is not allowed. This is shown in (31). In (31a), the clausal complement is in-situ, and the initial position of the matrix clause is filled in by a last resort excorporation of the lexical material of the inflected verb, leaving a semantically empty *do* auxiliary in Tense position. In (31b), the clausal complement fills in the initial matrix position, letting the lexical verbal material in the Tense position. The pattern illustrated in (31) is consistent across dialects.

(31) Standard (Kervella 1995[1947])

a. Karout a rafen [∅ e tiskouezfec'h an
love.INF FIN do.COND.PRES.1S C FIN show.COND.PRES.2PL the
dra-se dezhañ].
thing-here to.3SM
'I would like you to show this to him.'

b. * [∅ E tiskouezfec'h an dra-se dezhañ] a
C FIN how.COND.PRES.2PL the thing-here to.him FIN
garfen
like.COND.PRES.1S

However, in (32), MLB in Plougerneau (Leon) does allow for movement of a clausal complement (with an empty C)[9] in (32b). This clausal complement is compatible with stylistic fronting (c), but movement is ungrammatical if the clausal

[9] I leave unexplained why her grammar allows for this movement. Fronting of elements starting with a phonologically empty head is not usual.

complement is T2. The relevant minimal pair is between (32b) and (32d). Presumably, movement of the clausal complement must involve the complementizer, which forces it to linearize within the same domain as Fin prior to movement. This forced inclusion of C in turn prevents last resort operations for T2.

(32) Leon (Plougerne) (MLB 05/2018)

a. Tout an dud o d-eus keuz [∅ *e* *n-eus*
all the people FIN.3PL 3-AUX.PRES regret C FIN 3S-AUX.PRES
lennet Alain an dra-se].
read.PTCP Alain the thing-here
'Everyone regrets that Alain has read this.'

b. [∅ *E* *n-eus* lennet Alain an dra-se], tout an dud
C FIN 3S-AUX.PRES read.PTCP Alain the thing-here all the people
n-eus keuz.
(FIN) 3S-AUX.PRES regret
'Everyone regrets that Alain has read this.'

c. Tout an dud o d-eus keuz [∅ **lennet** *e*
all the people FIN.3PL 3-AUX.PRES regret C read.PTCP FIN
n-eus ~~lennet~~ Alain an dra-se].
3S-AUX.PRES ~~read.PTCP~~ Alain the thing-here
'Everyone regrets that Alain has read this.'

d. * [**Lennet** *e* *n-eus* ~~lennet~~ Alain an dra-se], tout an
read.PTCP FIN 3S-AUX.PRES ~~read.PTCP~~ Alain the thing-here all the
dud o d-eus keuz ∅.
people FIN.3PL 3-AUX.PRES regret C
'Everyone regrets that Alain has read this.'

4.3 Binding inside a T2 complement clauses

In the example (33), the quantifier 'nobody' is licensed by clausal negation. This quantifier binds the subject of the subordinate clause (33a). It does so equally when the past participle is positioned after the inflected auxiliary in a C-Fin-VSO structure, or before it, as is the case with stylistic fronting (33b).

(33) Leon (Plougerne) (MLB 05/2018)

a. Den ne lavar james [∅ *e* *n-eus* laeret an
nobody NEG say.PRES.3S never C FIN 3S-AUX.PRES steal.PTCP the
asiedoù kaer].
plate.PL beautiful
'Nobody ever tells he stole the beautiful plates.'

b. Den ne lavar james [∅ **laeret** *e* *n-eus*
nobody NEG say.PRES.3S never C steal.PTCP FIN 3S-AUX.PRES
~~laeret~~ an asiedoù kaer].
~~steal.PTCP~~ the plate.PL beautiful
'Nobody ever tells he stole the beautiful plates.'

In example (34), the subject of the upper clause, *Amelia*, c-commands the subject of the lower clause, *ar c'hruadur* 'the child'. The condition C prohibits the coreferencing of these two subjects, as evidenced by the impossible interpretation marked with an asterisk. Again, the element that underwent inversion through stylistic fronting is not the participle, but rather the subject. Consequently, the evidence is somewhat less robust, as it is challenging to identify contexts unequivocally precluding any form of focus fronting for this subject. Nevertheless, in each sentence, no explicit focus, such as contrastive focus, was discerned. Additionally, no distinct prosody was observed. For this neutral subject, its canonical position is immediately adjacent to the right of the inflected element, thereby rendering it a likely candidate for inversion.

(34) Leon (Plougerne) (MLB 05/2018)
Amelia e n-eus rentet kount [**ar c'huadur** *e-n*
Amelia FIN 3S-AUX.PRES give.PTCP report the child FIN-3S
n-eus kemeret an alc'hwezh fall].
3S-AUX.PRES take.PTCP the key wrong
'Amelia realized that the child had taken the wrong key.'
*'Amelia (=the child) realized that she had taken the wrong key.'

4.4 Non-deictic readings in temporal nominals

The final piece of evidence comes from the analysis of temporal nominal phrases. Phrases such as 'the current president' are typically constrained to deictic interpretations when used in root clauses. Consider, for instance, sentence (35a); the only plausible interpretation is that the clause's subject expresses dislike towards

someone who would subsequently become president, within a timeframe that encompasses the moment of utterance. Conversely, in clausal complement structures, temporal nominal phrases are not limited to deictic interpretations. Sentence (35b), a corpus example from Norland (2020), illustrates this point. Here, the interpretation is that the subject of the embedded clause has a dislike for the president who was in office at the time of this sentiment.

(35) a. He didn't like the current president.
b. He wanted to make a revolution for his country because he didn't like the current president.

In Breton, the temporal nominal phrase *va gwaz em beche* in sentence (36), which translates to 'my future husband', is not necessarily constrained to future time relative to the moment of utterance. This exemplifies that the second clause does not function as an independent root clause; instead, it is genuinely embedded within the syntactical structure, and interpreted as such.

(36) Leon (Plougerne) (MLB 05/2018)
Pa vedoun bihan, me soñje ∅ **va gwaz e-m**
when AUX.PAST.1S small I think.PAST.3S C my husband FIN-1S
b-eche *e veche* brasoc'h 'it ma 'z eo.
1.AUX.COND.PRES FIN 3.AUX.COND.PRES bigger than that FIN AUX.PRES.3S
'When I was a child, I thought my future husband would be bigger than he is.'

5 Linearization of high functional heads

The generalization presented in (37) captures the optional nature of T2 within embedded domains, and accounts for the observed compatibility of this optionality with last resort operations. Optionality is a result of the linearization of an initial functional head. When this initial head is not included in the linearization domain, a last resort operation becomes obligatory.

(37) Linearization can include or exclude an initial functional head iff it heads ForceP or a higher functional projection.

I leave open whether (37) can be derived by language-dependent prosodic rules. The dialect variation in the sporadic formation of T2 last resort embedded order arises from the dialectal parameterization of (37). All Breton speakers show a

preference for incorporating the complementizer within the spell-out domain of the embedded clause. In the Leon dialect, speakers like MLB and AM also accept configurations where this domain does not include the complementizer, thereby initiating last resort strategies. Conversely, Huguette Gaudart, representing the East Kerne dialect, consistently includes the complementizer in her linearization process, which means that for her, C-Fin-VSO orderings have no T2 alternative.

In the remainder of this section, I check (37) against functional heads other than ForceP. The sentence in (38) represents a polar question, commencing with an initial Q head that is followed by a complementizer, morphologically indistinguishable from a coordination marker. This complementizer may either be included into the linearization domain as in (38a), or omitted during linearization, a process which precipitates stylistic fronting as observed in (38b).

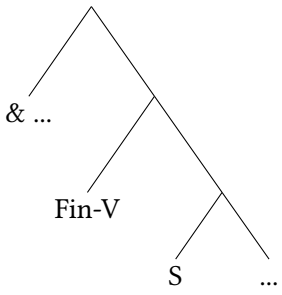
(38) Leon (Lesneven/Kerlouan) (AM 05/2016)

a. Daouste **hag** *e-n* *n-eus* diskouezet Anna laouen he loa,
Q if FIN-3SM 3-AUX.PRES show.PTCP Anna happy her spoon
pe inventet e-m eus?
or invent.PTCP FIN-1S AUX.PRES
'Did Anna proudly hold up the spoon or did I make it up?'

b. Daouste hag **diskouezet** *e-n* *n-eus* ~~diskouezet~~ Anna
Q if show.PTCP Fin-3SM 3-AUX.PRES ~~show.PTCP~~ Anna
laouen he loa, pe inventet e-m eus?
happy her spoon or invent.PTCP FIN-1S AUX.PRES
'Did Anna proudly hold up the spoon or did I make it up?'

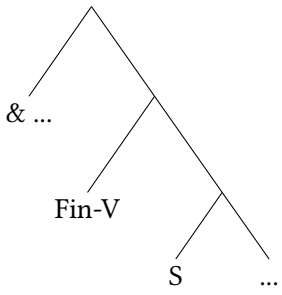
This effect can also be demonstrated using an actual coordination marker. Referring back to the structure of the left periphery as illustrated in (38), this marker is external to the clause, positioned above the C line. Demonstrating this effect with a coordination marker in a declarative clause proves somewhat more challenging. The issue does not lie in identifying the surface alternations schematized in (39). These alternations are fully productive. The coordination marker *ha*, which is pronounced *hag* before a vowel, can precede a regular T2 matrix sentence. This leads to &-T2 orders, as depicted in (39a), prior to a last resort operation for T2. An alternative, &-VSO orders, is also invariably feasible (39b), without triggering any last resort operation. In these instances, it is the coordination marker that occupies the initial position.

(39) a. & T2



& -X-***Fin-V*** S O ...

b. &-VSO



& -***Fin-V*** S O ...

The challenging aspect of our analysis involves demonstrating that no phonologically null element intervenes between the coordination marker and the Fin head in example (39b), and that indeed the initial element saturating the T2 requirement is the coordination marker and not a topic-drop or a null adverb.

Consider the sentence presented in (40). This sentence comes from a corpus.[10] The sentence preceding (40) in the context is unrelated, stating, "That was quite expensive at the time!". Hence, all information in (40) is contextually novel. This sentence likely exhibits a flat information structure with all-focus. The coordination marker introduces a temporal, consecutive interpretation. This interpretation complicates the task of conclusively ruling out the presence of a phonologically null adverb, such as 'then', which may have been omitted due to semantic redundancy with the coordination marker.[11]

(40) Leon (Plougerne) (Elégoët 1982: 16)
Hag *e* *koumañsemp* ar sezon da viz C'hwevrer betek fin
and FIN started.PAST.1PL the season at month February until end
miz Gwengolo.
month September
'And (then?) we used to start the work season in February until the end of September.'

However, certain instances exist where the coordination marker predominantly conveys an additive meaning – devoid of any temporal consecutive

[10] Specifically the transcription of an oral interview with a native Leon speaker from Plougerne, conducted in the 1980s.

[11] See Willis (1997) for the analysis of similar &-Fin-VSO orders in Middle Welsh, at the T2 stage of Welsh diachronic development. He discusses the possibility of an empty operator.

implications. In the preceding context for (41), the speaker stated "We had a long dock that we constructed ourselves, extending from the sea at its lowest point to the top. Imagine large rocks! Occasionally, they were as sizable as the table here…".

(41) Leon (Plougerne) (Elégoët 1982: 39)
Hag *e* *vije* m'en dare pet, gant limieri o klask
and FIN AUX.COND.3S 1S-3S know.not how.many with lever.PL at search
ober ar c'hal.
do.INF the dock
'And there was I don't know how many of them, with levers building the dock.'

I summarize in (42) the linearization patterns we have examined, for both matrix and embedded domains. The lowermost elements on the right side of the structure – focus, topic, or the functional negation head – mandatorily fulfill the T2 requirement, precluding any subsequent operations of a last resort nature. On the other hand, the uppermost elements on the left side, specifically the coordination marker, the Q head of polar questions and the C head of ForceP, demonstrate variability in their linearization with the verb. They obligatorily linearize with Fin in the Kerne dialect, and they preferably do so in Leon, with some sporadic exceptions that have no impact on semantics and show non-syntactic reflexes.

(42) & Q [hanging topics, scene-setting adverbs] C focus topic neg FIN-T …

Hanging topics and scene-setting adverbs, which are positioned between these functional high heads in the cartographic representation, never satisfy the T2 requirement. This characteristic correlates with their unique prosodic isolation, further supporting the argument that linearization plays a crucial role in this linguistic phenomenon.

Under the generalization that a coordination marker can be the first element in a T2 matrix sentence, Breton root-embedded asymmetries in word order is only epiphenomenal. Embedded domains are selected by complementizers, which are normally linearized in the initial position of the domain. Consequently, they are usually C-Fin-VSO. The same word order is observed in matrix sentences preceded by a coordination marker, & -Fin-VSO, adding to the list of matrix linear T2 orders illustrated from (4) to (7).

6 Conclusion

Building on previous research, I have proposed that the T2 requirement in Breton is a post-syntactic effect, an obligatory morphological exponence filtering word order at the scale of the sentence. I have rejected the hypothesis of Wurmbrand (2012; 2014) for Breton, and demonstrated that embedded T2 domains are fully integrated into the syntactic structure. The embedded T2 clauses under investigation are not syntactically root clauses. They are fully integrated at the syntactic level (cf. movement, binding), and are calculated as such by semantics (cf. readings of temporal nominal phrases).

The key element is that in the syntax of Breton, these embedded domains are not even T2. Like in any Celtic language, they are VSO. We saw that when an embedded clause is moved in syntax inside the domain of its matrix sentence, it is frozen to C-Fin-VSO order prior to this movement. Only in very special circumstances can the high complementizer be excluded from the linearization domain, calling for last resort strategies to ensure morphological exponence at the scale of the clause. The T2 filter is contained in another module operating after syntax, on the output of linearization. This is consistent with the diachronic evidence we have for T2 languages in Europe. The T2 property has a diachronic propensity for contagion by contact. The southern branch of Celtic languages has developed for centuries in contact with Old and Middle French T2. This contagion is efficient because it targets a morphological component (and probably prosody). Syntax is less prone to contact changes. After centuries of heavy Romance influence, Modern Breton for example still has its agreement system untouched, is still *pro*-drop, and has its subjects licensed post-verbally.

Linearization is subject, maybe through prosody, to dialectal parameterization. Breton dialects differ empirically in their tolerance for a leftmost functional head to be linearized separately. Only the Leon speakers allow it, even though they tend to dislike it. When they do so, they consequently deploy the last resort strategies available to them for Fin not to be linearized first. The impact of the interplay of linearization and prosody on Breton word order calls for further research. We need a proper modeling of the prosody of the sentence for each type of information structure that we detect in syntax and semantics, and especially the prosody of word orders that we know to be last resort solution for T2.

Appendix A High adjuncts and T2

Across T2 languages, including the Germanic domain, it has been noted that adjuncts vary with respect to the word orders they tolerate. Central adverbial

clauses resist T2, whereas peripheral adverbials do not (Haegeman 2012 and references therein).

A.1 Adjunct temporal clauses

In temporal clauses, we observe low adjuncts to be more restricted than high adjuncts. With the Leon speakers, diverse T2 orders are obtained with *pa* 'when', as we saw in (16), repeated here as (43) for convenience. stylistic fronting in (43b) is grammatical inside a temporal clause located in the recursive high left periphery, at the initial of the matrix clause, inside a prosodically isolated domain. In our analysis, stylistic fronting is a sign that the clause has been generated in this high position, as syntactic clausal movement would have frozen the structure to C-Fin-VSO orders.

(43) Leon (Plougerne) (MLB (05/2018)

a. **Pa** *m-eus* kleet ar janson-se evit ar wech kentañ,
when 1S-AUX hear.PTCP the song-there for the time first
eved-on dougeres.
C.AUX.PAST-1S pregnant
'When I first heard this song, I was pregnant.'

b. Pa **kleet** *m-eus* ~~kleet~~ ar janson-se evit ar wech
when hear.PTCP 1S-AUX ~~hear.PTCP~~ the song-there for the time
kentañ, eved-on dougeres.
first C.AUX.PAST-1S pregnant
'When I first heard this song, I was pregnant.'

In (44) and (45), low adverbials starting with the same complementizer *pa* are incompatible with stylistic fronting.

(44) a. Plijet e-n n-eus din **pa** *m-eus* klevet
please.PTCP FIN-3S 3-AUX.PRES.3S to.1S when 1S-AUX.PRES hear.PTCP
va mab e-n d-oa plantet va patatez.
my son FIN-3S 3-AUX.PAST plant.PTCP my potatoes
'I liked when I heard that my son had planted my potatoes.'

b. * Plijet e-n n-eus din pa **klevet** *m-eus*
please.PTCP FIN-3S 3-AUX.PRES.3S to.1S when hear.PTCP 1S-AUX.PRES
~~klevet~~ va mab e-n d-oa plantet va patatez.
~~hear.PTCP~~ my son FIN-3S 3-AUX.PAST plant.PTCP my potatoes
'I liked when I heard that my son had planted my potatoes.'

(45) Leon (Plougerne) (MLB (05/2018)

a. Amelia e n-eus c'hoarzet **pa** *n-eus*
Alemia FIN 3S-AUX.PRES laugh.PTCP when 3S-AUX.PRES
komprenet e n-oa en em dromplet.
understand.PTCP FIN 3S-AUX.PAST reflex mistake.PTCP
'Amelia laughed when she understood she had been mistaken.'

b. * Amelia e n-eus c'hoarzet pa **komprenet**
Amelia FIN 3S-AUX.PRES laugh.PTCP when understand.PTCP
n-eus ~~komprenet~~ e n-oa en em dromplet.
3S-AUX.PRES ~~understand.PTCP~~ FIN 3S-AUX.PAST reflex mistake.PTCP
'Amelia laughed when she understood she had been mistaken.'

A.2 Adjunct causal clauses

In causal clauses also, we observe low adjuncts to be more restricted than high adjuncts. The empirical observation is that negation in a matrix clause cannot scope over a T2 cause clause (see de Haan 2001, Heycock 2006, Haegeman 2012 and references therein; Holmberg 2009). This also holds true in Breton (46). The given context for (46a), 'This was a long trip', aims at suggesting the interpretation [¬ *he came*] *because he is lazy*. Structurally, it forces negation to restrict its scope to the matrix domain. When negation does not scope over the causal clause, but C-Fin-VSO orders (46a) and C-T2 are tolerated (46b).

(46) a. N' eo ket deuet **peogwir** *eo* lezireg, ha kavet
NEG AUX.3S NEG come.PTCP because AUX.3S lazy and find.PTCP
e-n d-oa hir an hent.
FIN-3S 3-AUX.PAST long the road
'He didn't come because he is lazy and he found it was too far away.'

b. N' eo ket deuet peogwir **lezireg** *eo* , ha kavet
NEG AUX.3S NEG come.PTCP because lazy aux.3S and find.PTCP
e-n d-oa hir an hent.
FIN-3S 3-AUX.PAST long the road
'He didn't come because he is lazy, and he found it was too far away.'

In (47), the given context is 'Don't be nasty! He didn't come with me only because I have a car, and he didn't want to walk…'. This is meant to force integration of the cause clause inside the scope of negation (¬ [*he came because he is lazy*]).

The embedded central adverbial clause is C-Fin-VSO in (47a) and excludes a T2 alternative (47b).

(47) a. N' eo ket deuet **peogwir** *eo* lezireg, met evit
NEG AUX.3S NEG come.PTCP because AUX.3S lazy but for
kaozeal samples.
discuss.INF together
'He didn't come because he is lazy but for us to talk.'

b. * N' eo ket deuet peogwir **lezireg** *eo* , met evit
NEG AUX.3S NEG come.PTCP because lazy AUX.3S but for
kaozeal samples.
discuss.INF together
'He didn't come because he is lazy, but for us to talk.'

The Breton facts align with the generalizations developed for Germanic languages. Peripheral adjuncts are attached high, outside of the scope of the negation in the matrix sentence. The high clause can be T2. Central adverbials however resist T2. The generalization is that the adjuncts which linearize in prosodically autonomous domains have the freedom to linearize with or without the initial complementizer. For some other reason, central adverbial domains resist this optionality. My proposal provides no new insight concerning this paradigm but suggests we should enrich the typological debate with prosodic information.

References

Bidese, Ermenegildo, Andrea Padovan & Alessandra Tomaselli. 2014. The syntax of subordination in Cimbrian and the rationale behind language contact. *Language Typology and Universals* 67(4). DOI: 10.1515/stuf-2014-0026.

Borsley, Robert D. & Andreas Kathol. 2000. Breton as a V2 language. *Linguistics* 38(4). 665–710. DOI: 10.1515/ling.2000.002.

Borsley, Robert D., Maria-Luisa Rivero & Janig Stephens. 1996. Long head movement in Breton. In Robert Borsley & Ian Roberts (eds.), *The syntax of the Celtic languages: A comparative perspective*, 53–74. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511586279.002.

de Haan, Germen. 2001. More is going on upstairs than downstairs: Embedded root phenomena in West Frisian. *The Journal of Comparative Germanic Linguistics* 1(4). 3–38.

Elégoët, Fañch. 1982. *Bezhinerien en enezeier: The seaweed seekers of the islands.* Lannion: Hor Yezh.

Haegeman, Liliane. 2012. *Adverbial clauses, main clause phenomena, and composition of the left periphery* (The Cartography of Syntactic Structures 8). Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199858774.001.0001.

Heycock, Caroline. 2006. Embedded root phenomena. In Martin Everaert & Henk C. van Riemsdijk (eds.), *The Blackwell companion to syntax*, vol. 5, 174–209. Oxford: Blackwell. DOI: 10.1002/9781118358733.wbsyncom068.

Holmberg, Anders. 2009. Verb-second. In Tibor Kiss & Artemis Alexiadou (eds.), *Syntax-Theory and analysis: An international handbook*, vol. 1. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110377408.342.

Jouitteau, Mélanie. 2010. A typology of V2 with regard to V1 and second position phenomena: An introduction to the V1/V2 volume. *Lingua* 120(2). 197–209. DOI: 10.1016/j.lingua.2008.11.011.

Jouitteau, Mélanie. 2011. Post-syntactic excorporation in Realizational Morphology: Breton analytic tenses. In Andrew Carnie (ed.), *Formal Approaches to Celtic Linguistics*, 115–142. Newcastle: Cambridge Scholars Press.

Jouitteau, Mélanie. 2012. Verb doubling in Breton and Gungbe: Obligatory exponence at the sentence level. In Enoch O. Aboh, Norval Smith & Anne Zribi-Hertz (eds.), *The morphosyntax of reiteration in Creole and non-Creole languages*, 135–174. John Benjamins. DOI: 10.1075/cll.43.04jou.

Jouitteau, Mélanie. 2020. Verb second and the left edge filling trigger. In Rebecca Woods & Sam Wolfe (eds.), *Rethinking V2*, 455–481. Oxford University Press. DOI: 10.1093/oso/9780198844303.003.0019.

Jouitteau, Mélanie. 2007-2024. *Arbres, site de recherche sur la syntaxe formelle de la microvariation syntaxique de la langue bretonne*. http://arbres.iker.cnrs.fr/ (5 June, 2025).

Kervella, Frañsez. 1995[1947]. *Yezhadur bras ar brezhoneg*. La Baule: 1947 édition: Skridoù Breizh, 1995 édition: Al Liamm.

Kiziak, Tanja. 2010. A controversial case; Extraction from V2 clause or parenthetical construction? In *Extraction asymmetries: Experimental evidence from German*, 81–122. Amsterdam: John Benjamins. DOI: 10.1075/la.163.

Lambert, Pierre-Yves. 1998. L'actance dans les langues celtiques. In Jack Feuillet (ed.), *Actance et valence dans les langues de l'Europe*, vol. 20 (Empirical Approaches to Language Typology 2), 295–346. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110804485.811.

Norland, Eric. 2020. *The wheels of friend: A three year around the world bicycle journey*. Bloomington: iUniverse.

Reis, Marga. 1997. Zum syntaktischen Status unselbständiger Verbzweit-Sätze. In Christa Dürscheid, Karl Heinz Ramers & Monika Schwarz (eds.), *Sprache im*

Fokus: Festschrift für Heinz Vater zum 65. Geburtstag, 121–144. Tübingen: Max Niemeyer Verlag.

Reis, Marga. 2002. Wh-movement and integrated parenthetical constructions. In C. Jan-Wouter Zwart & Werner Abrham (eds.), *Studies in comparative German syntax: Proceedings from the 15th workshop on comparative Germanic syntax*, 3–40. Amsterdam: John Benjamins. DOI: 10.1075/la.53.04rei.

Rezac, Milan. 2004. The EPP in Breton: An unvalued categorical feature. In Anne Breitbarth & Henk van Riemsdijk (eds.), *Triggers* (Studies in generative grammar 75), 451–492. Mouton de Gruyter. DOI: 10.1515/9783110197433.451.

Rivero, Maria-Luisa. 1999. Stylistic verb-movement in yes-no questions in Bulgarian and Breton. In Istvan Kenesei (ed.), *Crossing boundaries, advances in the theory of Central and Eastern European languages*, 67–90. Amsterdam: John Benjamins. DOI: 10.1075/cilt.182.06riv.

Rivero, Maria-Luisa. 2000. Finiteness and second position in long verb movement languages: Breton and Slavic. In Robert D. Borsley (ed.), *The nature and function of syntactic categories* (Syntax and Semantics), 295–323. New York: Academic Press. DOI: 10.1163/9781849500098_011.

Stephens, Janig. 1982. *Word order in Breton*. School of Oriental & African Studies, University of London. (Doctoral dissertation).

Travis, Lisa. 1984. *Parameters and effects of word order variation*. Cambridge: Massachusetts Institute of Technology. (Doctoral dissertation).

Willis, David. 1997. Clausal coordination and the loss of verb-second in Welsh. *Oxford Working Papers in Linguistics, Philology and Phonetics* 2. 151–72.

Wurmbrand, Susi. 2012. The timing of Merge: Deriving certain clause-linking mismatches. Paper presented at the Workshop on (Mis)matches in clause linkage. Leibniz Center for General Linguistics (ZAS), Berlin.

Wurmbrand, Susi. 2014. The merge condition: A syntactic approach to selection. In Peter Kosta, Steven L. Franks, Teodora Radeva-Bork & Lilia Schürcks (eds.), *Minimalism and beyond: Radicalizing the interfaces* (Language Faculty and Beyond 11), 130–166. Amsterdam: John Benjamins. DOI: 10.1075/lfab.11.06wur.

Chapter 3

Deriving VSO in Scottish Gaelic

Gillian Catriona Ramchand
UiT The Arctic University of Norway
Oxford University

In this chapter, I argue that standard approaches to word order typology in deriving the detailed properties of word order in Scottish Gaelic are too coarse and leave many properties unexplained. Data is drawn from adverb placement, copular clauses and light pronouns to highlight the empirical issues. I argue that a direct linearization approach to the mapping between syntactic representations and serial order is well motivated both empirically and architecturally. direct linearization contrasts with the family of approaches that places much of the typological burden on operations within the syntax itself in the form of "word order movements" (head movement, roll-up movements, EPP etc.). In the final sections of the chapter, I propose a novel direct linearization algorithm for Scottish Gaelic which is intended to open a theoretical conversation about what such a theory might look like, and how such theories might be explored.

1 Introduction

The leading intuition behind this chapter is that hierarchical representations are qualitatively different from the representations of linear order imposed by sensorimotor systems, and further, that syntactic implementations that build in a deterministic mapping between hierarchical or compositional relationships and sequencing are both descriptively and explanatorily inadequate. While the relationship between hierarchy and order is systematic *for a particular language*, and of necessity learnable, the relationship between the internal representations and their externalization is variable from language to language, and not one that falls out automatically from the shapes of the graphs we draw to represent the former. Berwick & Chomsky (2011) define externalization as:

Gillian Catriona Ramchand. 2025. Deriving VSO in Scottish Gaelic. In Andrew Carnie, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond & Luis Irizarry (eds.), *Foundational approaches to Celtic linguistics*, 97–139. Berlin: Language Science Press.
DOI: 10.5281/zenodo.15654861

> "Externalization – the translation of the hierarchical and recursive representations characteristic of syntax and semantics into the kinds of serial representations that the sensorimotor systems can manipulate" (Bennett et al. 2016b, on the Berwick & Chomsky 2011 notion of Externalization)

In this context, I take the order of words in a particular language to be a feature of externalization, not of hierarchical representation. Under this view, 'syntax' refers to representations of hierarchy, recursion and categorization; correspondingly, word order facts are not to be generated by the syntactic component of a grammar, but are evidence for the nature of an interface (in this case, language-specific though within broad constraints), since word order is just the output of a mapping from syntax to linear representation.

While most syntacticians would not reject the logical separation of hierarchy and word order, I would claim that working practice has not consistently operated with that intuition. We have been implicitly constrained by the graphical representations that we are accustomed to, and our analytic hypothesis space has been narrowed to favour solutions that embody universal interpretations of hierarchical representations in terms of ordering. In this exploratory and speculative chapter, I attempt to show that current approaches to word order generalizations in Scottish Gaelic (a VSO language) are deeply unsatisfactory. I propose a different implementation of the word order patterns that takes seriously the generation of word order as an output of a linearization algorithm without the intermediate step of displacement implemented in the syntax. I take the non-standard flavour of these exploratory solutions as evidence that they do *not* currently figure in our notional hypothesis space for these kinds of phenomena.

Typological work has long noted the systematic ordering differences found among the world's languages, and the implicational patterns that seem to be found there, suggesting some underlying system of universal constraints (Greenberg 1966). I begin by summarizing what I take to be the classical theoretical approaches to accounting for cross linguistic variation in base word order, as found in the literature.

Even with respect to the descriptive facts underlying something like Greenberg's Universal 20, one can broadly distinguish two major approaches:

(i) *Head ordering parameters*: Parametrize head-complement order directly (possibly on a category by category basis) (Haider 2000, Abels & Neeleman 2012 on Universal 20)[1]

[1]Base generation accounts do not eschew movement operations in the syntax entirely, but operate with a more restricted and triggered set of movement operations than Kayneian accounts.

(ii) *LCA-style universal mapping*: Assume a common phrase structural functional sequence, with a universal ordering (Kayne's LCA) and parametrize the choice of the kinds of movements that occur to distort base SVO phrase structures (Kayne 1994, Koopman & Szabolcsi 2000, Cinque 2004 on Universal 20).

To this, we can add a third option, which I call differences in direct linearization.

(iii) *direct linearization*: Assume a common relative hierarchical system across languages, plus a learnable linearization algorithm which is language specific.

The difference between direct linearization and the head-order parameter approach is that it is *not* assumed that the linearization algorithm is just confined to a parametric choice of head-complement order. The salient difference with LCA-style approaches is that it does not use any operations in the syntax designed purely to affect word order. This includes word order movements such as EPP movements, head movements, remnant movements, roll-up movements, etc.

2 Background on Scottish Gaelic word order

There is a respectable body of early (pre-LCA) work on deriving word order for VSO languages, including Scottish Gaelic and Irish. After the extremely common SVO and SOV languages, VSO is the next most common typological pattern with about 10 to 12 percent of attested world languages. When it comes to the head parameter however, VSO (and the less common VOS) are both considered subtypes of the head-initial setting for this parameter (Cinque 2017). Cinque even considers VOS to be the most pure, consistent word order type for the head-initial parameter setting, albeit noting that such 'consistency' seems to be cross-linguistically rare. In order to get the subject after the verb to derive the VSO preference from a base Spec-Head-Complement phrase structure, a number of different proposals have been made over the years, including a flat structure without a VP (McCloskey 1979, Chung 1983), lowering of the subject from a spec-IP position (Chung 1990, Shlonsky 1997 for Arabic), and raising of the verb to C as in V2 (Emonds 1980, Carnie et al. 2000 for Old Irish). The current consensus is that VSO languages like Scottish Gaelic and Irish do indeed have a VP, the subject remains in situ in a vP-internal or AspP internal position, and that the verb raises to some head position at the top of the inflectional zone higher than

the the subject (McCloskey 1990, 1996, Guilfoyle 1990, Noonan 1994, inter alia). See Carnie & Guilfoyle 2000 for further discussion.

If we take the embedded clause in Scottish Gaelic in (1) as an example, we see that the complementizer and the past tense particle precede the verb in the embedded clause, which is then immediately followed by the subject.

(1) Thuirt Calum [gun do leugh mi an leabhar].
say.PST Calum that read.PST.DEP(V) I(S) the book(O)
'Calum said that I read the book.'

This word order in the embedded clause is derived from a base generated SVO structure plus head movement of the verb to some position higher than the subject, as shown in the following tree. In this implementation, Fin, the lowest projection in the cartographic decomposition of the C layer is assumed to be the landing site for verb movement. The precise label for the head that is the landing site of verb movement here is not crucial; it merely needs to be a functional projection that is above the final position of the subject (here assumed to be Spec, TP) (Figure 1).

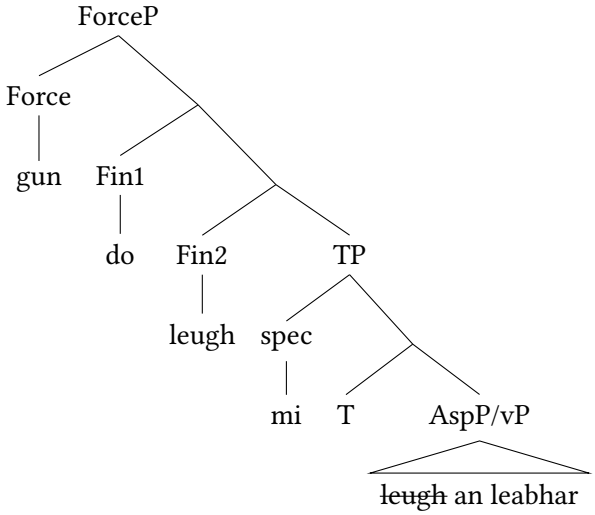


Figure 1: Tree for the embedded clause in (1)

The relevant research here comes from Irish (Carnie et al. 2000, Carnie 1995, McCloskey 1991, 1996, 2005, Noonan 1994). The standard account assumes further that there *is* a VP constituent headed by the verbal noun, and that the genitive

marked object also remains low within the VP. In some accounts, a higher functional projection is argued to host the direct case marked object in non-finite and perfect aspectual constructions. In Ramchand (1997) this is claimed to be AspP, while in Adger (1994) it is AgrOP since it co-occurs with agreement on the verbal noun. However, these details will be irrelevant for the problems I raise for the standard account below.

In addition to the position of the subject, there are a number of typological properties of the VSO/VOS family of languages that have been noted in the literature which set them apart from the other dominant head-initial type, SVO. Apart from the Greenbergian properties of postnominal adjectives, and the existence of prepositions rather than postpositions which go straightforwardly together with head-initiality, VSO languages all tend to have

(i) preverbal particles for higher functional elements in the verbal extended projection such as tense, mood, aspect, negation;

(ii) inflected prepositions (Kayne 1994);

(iii) left conjunct agreement (Doron 2000);

(iv) lack of a verb 'have' (Freeze & Georgopolous 2000);

(v) copular clauses without verbs (Carnie 1995);

(vi) "verbal noun" infinitives (Myhill 1985), as listed in Carnie & Guilfoyle (2000).

In addition, most VSO languages have VOS as a possible word order alongside VSO (Clemens & Polinsky 2013). At this point it is unclear whether, or how, these properties can be subsumed under a single property that unites VSO languages.

By all of the above criteria, Scottish Gaelic is a well behaved and typical VSO language, except for the possibility of VOS as an alternate word order for VSO order in regular clauses. I illustrate some of these basic word order properties below.

In (2) we see an example of a standard matrix clause with a transitive verb inflected for tense. The word order here is obligatorily VSO, with no VOS word order option available.

(2) Leugh mi an leabhar./(*Leugh an leabhar mi).
Read.PST I the book
'I read the book.'

In addition, Scottish Gaelic possesses auxiliary constructions in which the verbal information is split across a main verb carrying the event-descriptive content, and an auxiliary which bears tense and agreement (when it exists) information. In this kind of construction, it is the tensed auxiliary that is in the 'first' position, while the main verb (in the form a verbal noun – glossed as VN) stays low after the subject.

(3) Tha mi a' leughadh an leabhair.
be.PRS I Asp$_{ag}$ read.VN the book.GEN
'I am reading the book.'

While the position of subject and object always remains relatively fixed, objects either precede or follow the verbal noun in periphrastic constructions depending on aspect and finiteness: when they follow the verbal noun in the progressive aspect, they immediately follow it and are in genitive case (as we see in fact in (3). When they precede it as in infinitives or the perfect and prospective aspects, they take direct (Nom/Acc) case. An example of this for the perfect aspect is shown in (4) and for an infinitival clause in (5).

(4) Tha mi air an leabhar seo a leughadh.
be.PRS I Asp$_{air}$ the book this AGRO read.VN
'I have read this book.

(5) Bu thoil leam an leabhar seo a leughadh.
COP.PST pleasure with.me the book this AGRO read.VN
'I would like to read this book.'

Congruent with head-initiality, Gaelic has a wealth of prepositions, but no postpositions. It also has inflected prepositions, which is one of the typological properties listed above for VSO languages more generally. Given that this is rather unusual, in (6) I show the paradigm for the "inflected" preposition *le* 'with' as an example of this phenomenon.

(6) The preposition *le* 'with':

leam	'with me'
leat	'with you'
leis	'with him'
leatha	'with her'
leinn	'with us'
leibh	'with you.PL'
leotha	'with them'

With respect to the nominal extended projection, Scottish Gaelic is D/N initial, and has no pre-nominal modifiers apart from one or two special adjectives. It has neither pre-nominal relatives nor pre-nominal genitive phrases. The base order is therefore: Det N A Poss as shown in (7).

(7) taigh geal Chaluim
house white.M Calum.GEN
'Calum's white house'

The demonstrative consists of a post nominal phrase particle in combination with the initial determiner (8)

(8) an taigh seo
DET house DEM
'This house'

When it comes to *wh*-question formation, Gaelic again behaves as expected of a head-initial language, in always having the *wh*-pronoun (exemplified in (9) by *dé* 'what') in initial position.

(9) Dè a leugh thu a-raoir?
What COMP_{rel} read.PST you yesterday
'What did you read yesterday?'

In addition, conforming to the verb-initial property claimed by Freeze & Georgopolous (2000), Scottish Gaelic has no lexical verb 'have'. In (10), we see a sentence expressing possession. In Scottish Gaelic, this looks like a locative sentence using 'be', with the possessed element in subject position, and the possessor the complement of the preposition *ag* 'at'.

(10) Tha leabhar agam.
be.PRS book at.1S
'I have a book (The book is at-me).'

As already noted, Scottish Gaelic has no VOS alternative alongside VSO for ordinary verbs, but it is obligatorily subject-final for stative non-verbal predications involving the copula, as shown in (11) and (12). The correct phrase structure order for this construction is non trivial, given the final position of the subject, and the correct analysis is still controversial, but see Adger & Ramchand (2003) for a proposal.

(11) Is tidsear Calum.
COP.PRS teacher Calum
'Calum is a teacher.'

(12) Is leamsa an còta seo.
COP.PRS with.1S.EMPH the coat this
'This coat is mine.'

In short, Scottish Gaelic is a typologically "normal" member of the VSO family of languages by most diagnostics. This makes it plausible that any issues that arise in the discussion of how word order is generated for Gaelic will generalize to other languages with this basic profile. Having said that, this is not intended to be a chapter that examines the properties of VSO more generally, since it focuses empirically only on Gaelic (with some reference to established facts concerning the closely related Irish (also Goidelic) language. I argue that there are serious descriptive problems in characterizing the word order of Gaelic within a standard set of syntactic assumptions, even though these languages are classically considered to be one of the success stories of the head movement approach to verb placement.

Summarizing the consensus briefly once again, it is often claimed for Scottish Gaelic that there is a VP that contains both the subject and the object, and that the tensed verb moves to a higher functional head in the inflectional domain to give verb-initial order. Higher functional positions host particles such as negation and complementizers which precede the lexical verb forming an integrated prosodic verbal complex. Auxiliaries base generated in the inflectional zone are also attracted to the high inflectional position, while the verbal noun spells out the verbal position within the VP. Functional material in the middle inflectional zone such as Asp or Agr have been proposed to host preposed objects in infinitival and perfect aspectual constructions.

Within a framework that assumes the LCA, VOS/VSO is considered to be one of the most consistent versions of a descriptively head-initial parameter setting which regularly produces word orders that orders phrasal elements to the right of the head in reverse hierarchical order, and functional elements before the lexical head in straight LCA compliant order (Cinque 2017). In the LCA compliant approach, this is delivered by roll-up movements of phrases including V involving pied piping of the *whose*-pictures type (detailed derivations given in Cinque 2017). My position in this chapter is to reject a framework that involves word order movements and explore a direct linearization approach, so I will not detail those derivations here. However, it is worth pointing out that such derivations, if

implemented consistently for Scottish Gaelic would predict at least the possibility of ordering the direct object before the subject in a standard transitive clause. While, as we will see, adverbial modifiers do indeed stack up in reverse hierarchical order in Gaelic, the same is not true of arguments. This means that a principled distinction needs to be made between arguments and phrasal modifiers in the phrase structure in order to even state the pattern. In many cartographic approaches, null functional heads are assumed to host phrasal modifiers in their specifiers (Cinque 1999), making them phrase structurally indistinguishable from A-positions which host arguments. An algorithm which generates movements based on a certain kind of pied-piping will affect both arguments and adjuncts equally unless a principled distinction is made between functional heads of a "core" projection that introduce arguments and distinct functional heads introducing adjuncts. Something like this is what Cinque (2017) suggests. At any rate, it is not the case that the problems for the head ordering parameter account can be solved by embracing the LCA and word order movements.

While the assumptions (head-initiality, or its LCA equivalent, plus head movement) go a long way towards describing most of the major clause types of Gaelic, there remain a few knotty problems and word order patterns that are more difficult to account for with this simple set of parametric choices.

In the section that follows, I lay out some empirical challenges for the standard accounts. The first set of problems involves the placement and scope of adverbials, with some differences for simple vs. auxiliary verb constructions. The second problem involves copular clauses. The third problem involves the placement of light object pronouns (a problem already discussed by Bennett et al. 2016a for Irish). The purpose of this section is to emphasize that word order has not yet been "solved" for any of the non Spec-Head-Complement languages, let alone for Gaelic. It is not simply a matter of the linguist being forced to choose between two different kinds of implementations with slightly different universalist implications; the patterns here are going to require extra or different mechanisms to state. They therefore invite us to rethink the architectural model and the relationship between syntax and linearization, and to experiment with different ways of stating the generalizations we find.

(13) Problems:
   a. Extra stipulations are required to ensure that none of the specifiers of those higher heads can be filled, plus ad hoc prohibitions on adjunction.
   b. Copula clause word order.
   c. Weak pronoun placement.

3 Problems for the standard account

While the set of assumptions discussed in Section 2 correctly predicts the height of the verb with respect to subject and object in main and subordinate clauses, it has nothing to say about the positioning of adverbials and other adjoined phrases unless we subscribe to a view where adverbials are rigidly and templatically ordered. Scottish Gaelic is discernibly different from both V2 languages and language such as English, when it comes to the possible positions of adverbials. Let us take a broad look at the data to see whether the templatic account, combined with the "high" position of the verb makes the right predictions.

3.1 Adverb placement

For the purpose of illustration, I concentrate on the two broad classes of adverbials established in the literature VP-adverbials and sentential adverbials which have different syntactic position and semantic scope (Bellert 1977, Jackendoff 1972).[2] Adverbials that describe the manner of the action of the verb can be adjoined to the VP in English, while evaluative or speaker oriented adverbs tend to take the whole sentence in their scope and semantically modify whole propositions. This gives rise to the asymmetry in ordering positions shown in (14).

(14) a. Fortunately, John climbed the tree carefully.
b. * Carefully, John climbed the tree fortunately.

While a sentence adverbial can be left peripheral when the manner adverbial is right peripheral, the opposite does not sound very good in English. In addition, English seems to have a third possible position for adverbials in between the subject and the tensed verb. Subject-oriented adverbs such as *cleverly* seem particularly felicitous in this position, although all types of adverbs seem to be able to appear linearly in here with the right intonation. In general, however, there seems to be a hierarchical ordering of semantic adverb types that finds cross-linguistic support. In the following hierarchy, I confine myself just to the three basic old fashioned types, ignoring the more articulated proposals of e.g. Cinque (1999).

[2] Of course, in the theoretical literature since the seventies many more types of adverbials have been proposed together with finer grained structural positioning and syntactic scope (Cinque 1999). However, the more subtle the distinctions, the more difficult it is to establish clear ordering judgments. I concentrate on two broad syn/sem classes here because the problem can be more robustly demonstrated with them.

(15) Speaker oriented adverbs < Subject oriented adverbs < Manner adverbs. (Jackendoff 1972, Alexiadou 1997, Sportiche 1988).

If we take a manner adverbial[3] such as *'s a'bhad*, 'quickly' in Scottish Gaelic, we find that the most natural position for it is in fact clause finally, as shown in (16).

(16) Dh'ith Iain an ubhal 's a'bhad.
eat.PST Iain the.FSG apple.FSG quickly
'Iain ate the apple quickly.'

While it is possible to place a manner adverb in a clause initial position in English, with suitable comma intonation, as shown in (17).

(17) Quickly, Iain ate the apple.

The equivalent ordering is unacceptable in Scottish Gaelic.

(18) a. * 'S a'bhad dh'ith Iain an ubhal.
b. * Dh'ith 's a'bhad Iain an ubhal.
c. * Dh'ith Iain 's a'bhad an ubhal.

This is an instance in which Gaelic's preference for verb initiality seems to go beyond a simple requirement that the verb be ordered before the subject. However, if we change the adverb to a speaker-oriented evaluative sentence adverb, then the sentence initial position does become felicitous in Scottish Gaelic, with what sounds like two separate phonological phrases being formed. The sentence final position is also grammatical for subject-oriented adverbs. Finally, for speaker-oriented (high) adverbials, we find that the initial position is possible, with appropriate intonation (19), while the sentence final position is once again quite natural (20).

(19) *Gu mi-fhortanach,* dh'ith Iain an ubhal.
unfortunately eat.PST Iain the.FSG apple.FSG
'Unfortunately, John ate the apple.'

[3]The following adverbial data from Scottish Gaelic comes courtesy the University of Arizona archive of Scottish Gaelic data, and comes from the results of a number of targeted questionnaires on clausal ordering. Thanks to Andrew Carnie (PI) and Muriel Fisher (primary consultant) for this resource.

(20) Dh'ith Iain an ubhal *gu mi-fhortanach.*
eat.PST Iain the.FSG apple.FSG unfortunately
'John ate the apple, unfortunately.'

One very strong constraint is that in Scottish Gaelic, it is not possible to place an adverb in between the subject and the direct object, regardless of whether that adverb is a manner adverb as in (21), or a speaker oriented adverb as in (22).

(21) * Dh'ith Iain *'s a bhad* an ubhal.
eat.PST Iain quickly the.FSG apple.FSG

(22) * Dh'ith Iain *gu mi-fhortanach* an ubhal.
eat.PST Iain unfortunately the.FSG apple.FSG

Interestingly, when the two types of adverbs combine we get an asymmetry. The "low" VP adverbial needs to be expressed before the the "high" one, in a left to right order that tracks the cross-linguistically attested hierarchy, as shown in (23).

(23) Dh'ith Iain an ubhal *'s a bhad*, *gu mi-fhortanach*
Eat.PST Iain the apple quickly unfortunately
'Iain ate the apple quickly, unfortunately.'

As we have noted previously, the reverse hierarchical order of adjuncts is a common feature of many head initial languages and is usually handled via roll-up movements and *whose*-pictures pied-piping.

So far, the adverbs have been sentence peripheral, but there is a class of construction types where we do find the possibility of a sentence-medial adverbial position. This occurs when you have an auxiliary construction, with the tensed verbal element being distinct from the contentful verb expressed as a non-finite verbal noun/participle.

(24) Bhiodh bana-bhuidsich *gu tric/gu slaodach* a' briseadh nan
be.COND witches often/slowly PROG break.VN the
sguaban aca.
broomsticks at.3PL
'Witches would often/slowly break their broomsticks.'

The position in between the subject and the aspectual phrase is possible for these adverbials but no adverbials seem to be possible in the stretch between the aspectual particle and the verbal noun, or between a verbal noun and its direct object in the genitive.

(25) Tha oileanach (*gu tric*) air (**gu tric*) na leabhraichean sin
be.PRS students (often) PERF (often) det.DIR.PL book.PL those
(**gu tric*) a reic *gu tric* ann a shin.
(often) PRT sell.VN (often) there
'Students have often sold books there.' (from Adger 2010)

The position between the tensed verb and the subject is also quite restricted. There are however some constructions where it has been argued that the grammatical subject stays low within the VP (for example, in existential constructions), and in those cases, a suitably scoping adverb is possible before the subject (26).

(26) Tha *gu tric* oileanaich *gu tric* 'san leabharlann.
be.PRS often students often in.the library
'Students are often in the library.'

In contrast, the classical subject position (the specifier of TP) cannot be separated from the tensed element by adverbs or by parentheticals.

(27) Tha (**gu tric*) Calum *gu tric* 'san leabharlann
be.PRS (often) Calum often in.the library
'Calum is often in the library.'

(28) Tha (**tha mi cinnteach*) Calum *gu tric* 'san leabharlann
be.PRS (I am sure) Calum often in.the library
'Calum is often in the library.'

For Irish, Elfner (2015) shows that in sentences such as (29), there is little or no evidence for a prosodic boundary between the Verb and the Subject, while there is strong evidence for a prosodic boundary between the Subject and the Object.

(29) Chonaic Máire Seán.
Saw Mary John
'Mary saw John.' (Irish)

The facts about adverbial placement in Scottish Gaelic summarized here seem to give evidence for hierarchy, since different scopal classes of adverbs have different adjunction possibilities. Mid and high adverbs like *gu mi-fhortanach* 'unfortunately' can be both initial and final, while low (manner) adverbials can only be final in a simple clause. When these classes of adverbs co-occur in final

position, they occur in reverse hierarchical order with manner adverbs preceding subject oriented, preceding speaker oriented adverbials. In medial position, which is only possible with periphrastic (auxiliary) constructions, both mid and low adverbials can occur between the subject and the aspectual particle.

Even though there are clear patterns here, it is not clear how to model them without stipulating in some way *which* projections with which directionality are allowed for adverbial adjunction/modification, or how to allow optionality in the absence of semantic effects. The head-parameter (or its implemented equivalent in the type of pied piping required in roll-up movements) is also not well suited to making a difference between adjuncts and arguments if they are both phrases in specifier positions, or in preventing certain classes of adverbials from e.g., adjoining to the 'left' of the VP they modify, as is the case here with manner adverbials in Gaelic. Quite generally, it is not considered to be the job of word order typological theory to regulate the choices and preferences for elements like adjuncts that have a freer distribution. Yet, we do find language specific preferences and requirements in this domain as well. Going back to the difference between English and Scottish Gaelic, we find that Gaelic does not generally allow adverbial elements to stack up at the beginning of a sentence. In English, the following sentence in (30) sounds quite natural.

(30) Yesterday, in the bathroom, with great reluctance, John shaved his mustache off.

The equivalent in Scottish Gaelic would sound very awkward if not completely ungrammatical (31).

(31) ?? An dè, anns an seòmar-ionnlaid, gu mòr leisg, chuir
Yesterday in the bathroom with great reluctance put.PST
Iain dheth an mustache aige.
Iain of.him the mustache at.him

The preferred order of adjuncts here, would be for all of them to sit in the sentence final position, in reverse order compared to the English one (32).

(32) Chuir Iain dheth an mustache aige gu mò leisg, anns an
Put Iain off.him the mustache at.3SM with great reluctance in the
seòmar-ionnlaid, an dè.
bathroom yesterday
'Yesterday, in the bathroom, with great reluctance, Iain shaved his mustache off.'

In some sense, we could see this preference in Scottish Gaelic as an extension of its verb initial property, since the dislike of sentence initial adverbial phrases is one of the things that delivers the overwhelming preponderance of pure verb initial sentences in the language.

As mentioned earlier, some recent cartographic accounts propose to directly explain adverb ordering by interpreting adjunction as the filling of a specifier of particular functional heads. Under this view, the adjunct is semantically related to the content of the assumed functional head (which may be non-overt), and moreover, the relevant functional heads are hierarchically added to the core lexical phrase in a particular deterministic (often assumed to be universal) order (Cinque 1999). This way of approaching the problem gives rise to the expectation that adverbs will appear in the same relative order in all languages, and that this will appear in lockstep with the heads that represent those particular abstract meanings. It also means that a parametric setting for head initiality (a certain kind of pied-piping during roll-up) should treat arguments and adjuncts alike within a particular language. Argument positioning in Scottish Gaelic is quite distinct from, and significantly more rigid with respect to head positions than adverbial positioning is, and the relative order of subject and object is always left to right respecting hierarchical order. Adverbials on the other hand stack up in *reverse* hierarchical order when they follow the verb, and left 'adjunction' seems rather restricted. If we decide on the other hand that adverbs do attach in a syntactically distinct way to arguments in specifier positions via adjunction (Roberts 1996, Kayne 1994), then word order patterns built around arranging the linear order of head-complement or specifier-head do not necessarily bear on adverbial placement and ordering at all. We have already seen that there *are* generalizations about relative adverb ordering across languages, so relating these adjunct positions to their scopal height, or height of certain functional material seems like a promising strategy (cf. Ramchand & Svenonius 2014 for a zonal account) but doesn't solve the adverbial ordering problem directly or explain why it is not exactly the same as the ordering of heads and arguments in some languages.

Moreover, within Gaelic, we have at least one hard empirical problem that militates against the functional projections solution to adverb ordering. This is the issue of medial adverbials. As we have seen, in sentences with a simple tensed main verb, it is not possible to insert adverbials in between the subject and the object. However, if the construction is periphrastic, this suddenly becomes a possibility. (The minimally contrasting sentences are shown below).

(33) Bhiodh bana-bhuidsich *gu tric/gu slaodach* a' briseadh nan
be.COND witches often/slowly PROG break.VN the
sguaban aca.
broomsticks at.3PL
'Witches would often/slowly break their broomsticks.'

(34) Bhris a'bhana-buidseach **gu tric/*gu slaodach* nan sguaban aice.
be.PST the witch often/slowly the broomsticks at.3SM.
'The witch often/slowly broke her broomsticks.'

The problem this raises for cartographic approaches to adverb ordering is the following. We would like to assume that the functional sequence for periphrastic and synthetic verb phrases is exactly the same within a single language, since the actual spell out of the syntactic structure should not influence this (under standard assumptions of late insertion). This means that the two sentences above should have exactly the same abstract projections ordered in the same way. It is unclear, within this framework, how one would forbid the merge and left-linearization of an adverbial in the one and not the other.

This long digression into adverbial placement preferences in Gaelic has been to emphasize that accounting for language particular word order facts is not confined to regulating the order of head, complement and specifier. If adjunct placement is taken into account, a host of other generalizations emerge that have to be independently stated in some way. In the final section of this contribution where I introduce my own speculative algorithms, the goal will be to bring these sorts of facts into the fold as parts of the explanandum for linearization algorithms.

To summarize this section, getting V-initial order as a more general output crucially relies on preventing adverbial attachment to the left of the projection that the verb raises to. The preference for right adjunction of adverbials in Scottish Gaelic needs to be stipulated and doesn't currently follow from anything, but it is arguably part of the conspiracy that delivers V1 in the first place. Moreover, Scottish Gaelic seems to be showing us that a principled distinction needs to be made between the ordering of arguments (assumed to be filling specifier positions of heads) and the ordering of adjuncts. Finally, under a universalist phrase structure template it is hard to see how to derive the different possibilities for medial adverbs just when the phrase structure is filled out in a periphrastic way as opposed to a "synthetic" way.

3.2 Copular clauses

In this subsection, I describe a word order pattern within Gaelic that has proved recalcitrant over the years, if a general theory of phrase structure is to be proposed for the language. The clause type I am referring to is the copular clause, shown below in (35). In (35), we see the tensed copula *Is*, followed by the adjectival predicate *mor* 'big', while the subject of the sentence *an duine sin* 'that person' is sentence final.

(35) Is mór an duine sin.
COP.PRS big the person that
'That person is big.'

This pattern for copular clauses is quite general, regardless of the syntactic category of the predicate. The copula and predicate phrase form a unit to the exclusion of the grammatical subject, which is final. In (36) , we see a prepositional phrase in predicate position, showing that head movement to the high copula position cannot be the derivation that delivers this word order pattern.

(36) Is le Calum an cù.
COP.PRS with Calum the dog
'The dog is Calum's.'

In (37), we see a noun phrase in predicate position, once again showing the categorial generality of the structure and the phrasal nature of the element that immediately follows the tensed copula.

(37) Is toil leam an cù.
COP.PRS liking with.1S the dog
'I like the dog.'

This has had many different treatments in the literature, but the movements that would derive this order from a sensible hierarchical base structure seem to violate certain basic laws of syntactic movement, as commonly construed. For example, in Carnie (1995), complex predicational phrases must be allowed to "head move" in these constructions. That the final DP is the subject of the copular predication can be shown by the following datum on reciprocal binding in (38) from Irish (Doherty 1996).

(38) Is cosúil lena chéile iad.
COP.PRS like with.3PL.POSS each.other they
'They are similar to each other.' (Irish)

The standard phrase structure for declarative clauses in Gaelic has a high T position which is usually filled by an auxiliary or by head movement of a main verb from lower in the structure. In the copular construction, we find the same tense-initial position, but the equivalent of the VP looks as if it has raised past the subject to be adjacent to this tensed position. Although these constructions do not involve transitive predicates with direct objects (the copula combines with APs, PPs and NPs), the order here is reminiscent of the VOS orders often found in verb initial languages, in having the predicate phrase before the subject.

Once again, this is not the sort of word order problem that traditionally falls under the remit of word order typology, or word order parameters. And perhaps it is a mistake to consider this construction type at all when characterizing the Gaelic linearization algorithm, although *some* way of representing it phrase structurally needs to be found that fits our expectations of hierarchy and ultimate word order. If this order can be made to fall out of a more general way of thinking about how to characterize Scottish Gaelic word order, it would be a case of solving a problem that does not currently figure in any of the other implementations for Scottish Gaelic syntax to word order mappings.

3.3 The light pronoun problem

Finally, I turn to yet another word order issue that has not traditionally been considered part of the word order typology. This is a phenomenon found in Scottish Gaelic and Irish completely independently of VSO, that raises severe problems for the relationship between hierarchical representations and linear order. I include it here because it is a flagship case for more direct approaches to linearization (i.e. ones that do not avail themselves of an intermediate stage of syntactic displacement). Since my own account of word order facts more generally is going to be in the direct linearization family, it will be relevant to see this last problem too in the context of the more general proposal for delivering Verb initiality.

The light pronoun problem concerns the linearization of direct object pronouns within the VP. To see the nature of the pattern, we need to first look again at where object nominal phrases are linearized in a standard Scottish Gaelic sentence. As we can see in (39) below, the direct object *Iain* comes right after the grammatical subject, and before the two adjunct adverbials.

(39) Chunnaic mi Iain 's ghárradh a-raoir
saw I Iain in.DEF garden last night
'I saw Iain in the garden last night.'

(40) Chunnaic mi Iain a-raoir 's ghárradh
saw I Iain last night in.DEF garden
'I saw Iain in the garden last night.'

As we saw in the section on adverbials, in a sentence such as this, the direct object always precedes both adverbials, and although there is some flexibility in the order of those adverbials, the preference is for the higher scoping adverb to be rightmost.

The situation changes however, when the direct object is expressed as a light pronoun. While the order that replicates the order found with full nominal containing phrases is fully grammatical, as shown in (41).

(41) Chunnaic mi e 's ghárradh a-raoir
saw I him in.DEF garden last night
'I saw him in the garden last night.'

Two additional word orders emerge as natural here, namely the positioning of the light object pronoun after adjunct 1, or adjunct 2 (42).

(42) a. Chunnaic mi 's ghárradh e a-raoir
b. Chunnaic mi 's ghárradh a-raoir e

These linearization options are found only with 'light' object pronouns and not for object phrases more generally. It can also be shown to carry no consequences for information structure (Bennett et al. 2016a).

One could imagine trying to use syntactic movement to account for pronoun placement (like e.g. object shift in Scandinavian), but the rightward nature of the movement and the restriction to light pronominal elements is a first indication that the precise solution is likely to be complex. Movement accounts have indeed been explored, but they have been shown to have deep, potentially fatal problems. In particular, Bennett et al. (2016a) and Adger (2007) show that if this variability in linearization possibilities were to be captured via a syntactic movement, that movement would have to have the following undesirable properties.

(43) a. It would have to be able to lower into islands.
b. It would not feed other syntactic movements.
c. It would not feed ellipsis.
d. It has no information structure import.

Taking stock briefly of the problems for word order theory I have discussed in this section, I think that to achieve descriptive adequacy beyond the coarse-grained Greenbergian classes a different kind of theory is required. The empirical challenge includes characterizing the adjunction possibilities of modifiers as distinct from arguments, as well as the positioning of rogue elements such as light pronouns and lexically driven constructional idiosyncrasies, such as we have seen in Scottish Gaelic. In short, the problem of how to characterize the linearization properties on a language specific basis of a given hierarchical representation has not really been solved.

4 Removing word order from the syntax

We have seen that one dominant approach to characterizing the mapping from structure to serialization is to stipulate that the mapping is rigid and universal but then parametrize the different movements within the syntactic component that operate to deform the original hierarchy. In all these cases, the original hierarchy is well motivated from a typological point of view, reflecting the constituents and embedding relationships that also make sense from a syntactic diagnostic and semantic perspective. The final deformed structure created by word order movements (especially remnant and snowballing movements) do not on the other hand retain any kind of internal logic or intuitive constituency but are there solely to deliver the correct serialization based on the LCA. Given the fact that these movements are also rather different from those standardly allowed in the syntactic component, an obvious conclusion to draw is that implementing the mapping between hierarchy and serialization as a topological re-morphing within the syntax itself is a kind of category mistake. We *know* that language is hierarchically structured, and we have ample evidence for distinct parts of speech, selectional relationships, agreement relationships etc. both descriptively and within processing. The syntax component is rich and internally structured with its own well studied primitives and relations (including a class of well motivated 'displacement' relationships). We also know that languages are communicated serially as part of the externalization process and that different languages serialize 'equivalent' syn/sem representations differently. The minimal, first-principles strategy should be to investigate directly how these two different independent and justifiable representations (the hierarchical and the serial) map on to each other and what patterns and generalizations concerning that mapping emerge across languages. The dominant, movement strategy turns that mapping into a part of syntax itself, a kind of intricate distortion algorithm which,

while in principle could lead to observationally accurate output, creates a kind of middle man which neither helps us understand the nature of syntax itself better nor is helpful in connecting to the interfaces with cognition more generally.

In this section, I summarize some recent work to argue that we have good reasons to believe that linearization algorithms are part of a domain general translation between hierarchy and serialization, and not a parochial syntax-internal set of mechanisms. That some linearizations of a hierarchy are possible mappings while others are not is an important fact and one that we would like to understand better. Some recent work that doesn't filter the generalization through the language of movements, offers an intriguing possibility for where the explanation might lie. Medeiros (2024, 2018) makes the important observation that if you have a hierarchy (stated as a sequence 1,2,3,4 with 1 the most deeply embedded), then the typologically attested orders are just those that can be sorted into that hierarchy by Knuth's simple stack-sorting algorithm (Knuth 1968). In the statement of the algorithm below, imagine an input sequence with first incoming element I, a separate stack with top element S fed by the Push operation which places an element from the input on the top of the stack, and a final output that is fed by the Pop operation from the top of the stack to the output.

(44) Stack-sorting algorithm (from Medeiros 2024)
While input is nonempty,
If S < I, Pop
Else, Push.
While stack is nonempty,
Pop

Essentially, if the input element is lower in the hierarchy than the element on the top of the stack then it is pushed onto the stack, becoming the top element itself. If the input element is higher in the hierarchy than the top of the stack then it is sent immediately to the output. For example, let us assume we are trying to output the base hierarchical order 123 (where 1 is the most deeply embedded) then if the *input* is 123, identical to the desired output, then each element is pushed and immediately popped. On the other hand, if the input is the mirror image of the output, 321, the entire sequence is first pushed onto the stack, and then popped one by one, reversing its order. The stack essentially acts as a buffer, keeping track of the order of incoming elements and keeping that in "memory" while/if the elements from the input are still decreasing in functional height. An increase in functional height from the input triggers a pop of the lower element from the stack into the output. In this way, this simple algorithm takes a serialized input

of elements and can convert it to a sequence that represents hierarchy, in an order suitable to feed bottom up compositional interpretation. However, not all possible orderings of a 1-2-3 base hierarchy can be stack-sorted back into 1-2-3 by this simple algorithm. There are six possible permutations of 123, five of which will successfully return 1-2-3 ((123, 132, 213, 321, 312), and one of which will not (231).

	Output	Stack	**I**nput
	1-2-3 ⟵	⋃	⟵ 1-2-3
	1-2-3 ⟵	⋃	⟵ 3-2-1
(45)	1-2-3 ⟵	⋃	⟵ 1-3-2
	1-2-3 ⟵	⋃	⟵ 2-1-3
	1-2-3 ⟵	⋃	⟵ 3-1-2
	*2-1-3 ⟵	⋃	⟵ 2-3-1

The typological generalizations underlying Greenberg's Universal 20 have recently proved a fertile testing ground for theories of word order parametrization and the balance between head ordering parameters and different types of movements (Cinque 2017, Abels & Neeleman 2009). Strikingly, when it comes to Universal 20 in particular, Medeiros (2024) demonstrates how the orders that are "unsortable" by stack sorting are precisely those which are claimed to be unattested in a typological survey of NP internal orderings of Dem, Num, Adj N. Medeiros (2024) also argues that the Final-Over-Final Constraint (FOFC Biberauer et al. 2007) can also be derived from a prohibition against non stack-sortable orders.

I follow the general assumption that some abstract hierarchical ordering of elements is linguistically universal and that this hierarchical arrangement of symbols is crucial to meaning composition in human languages (while remaining agnostic for the time being about whether languages all need to agree on the number of fine grained distinctions and functional elaborations in each domain they choose to grammaticalize). If this is the source of the hierarchical representations in all languages, then these base representations should be the same for all languages, modulo differences in fine-grainedness and in which of the hierarchical elements is grammaticalized. Different word orders for the same elements must therefore emerge from the language particular mapping between those structures and a linearized representation suitable for externalization.

What does this mean? Knuth's stack sorting algorithm (and Medeiros' implementation of it) is not some sort of mystical incantation. It is an algorithm designed to do one functionally important thing – from an input *sequence*, it transforms the material to output elements in the order of bottom to top hierarchical embedding. This simple algorithm can handle nearly any choice a language

might come up with for how the linear order of symbols might proceed and fails on only a few (from the point of view of the algorithm) "pathological" choices. It turns out that human language ends up avoiding these pathological choices. This suggests that something in the way that language is set up *makes* those order choices pathological, suggesting in turn that the stack sorting algorithm shares logical properties with how human speakers actually parse incoming sequences as well.[4] In other words, the attested word orders are the ones that one can use a simple domain general "processing" algorithm for to recoup the underlying hierarchy.[5]

What I take this to suggest for our discussion of ordering is that the cluster of constraints on movement necessary to derive the typological pattern of *2-3-1 are plausibly not a syntax-specific set of constraints on movement, but reflect a more general informatic principle concerning how one recoups hierarchical dependencies from a linear presentation. Basically, it shows us that the typological generalizations underlying the modern understanding of Greenberg's Universal 20 should not be seen as arguments for a parametrization for movement plus *whose*-pictures type pied piping vs. pictures-of-*whom* type pied piping in syntax. Rather, they say something about the way working memory works to support the representational conversion from strings to underlying hierarchy. Seeing the algorithm from a direct linearization perspective allowed us to see it more abstractly in terms of working memory and tracking of selection. On the other hand, movement algorithms are so parochial to the vocabulary of syntax that they obscure this more abstract interpretation. This is one example of how a difference in implementational choice is not mechanistically innocent, even when the two versions deliver the same results observationally.

As fascinating as the stack-sorting generalization is, it still does not actually answer our questions about the *particular* stack-sorting compliant word order a language ends up with, nor does the stack sorting have anything immediately

[4] I am assuming here that the constraint comes from the mechanisms of comprehension rather than production, since this algorithm sorts left-right sequences into bottom-up representations that would naturally feed a comprehension process. Abels (2016) points out that if you wanted stack-sorting to fail on the typologically bad orders when considering the *production*-motivated mapping from hierarchy into order, then the input would have to be top-down and the output right-to-left to get the correct typological pattern.

[5] Interpreting the stack-sorting generalization in this (possibly simplistic) processing way, assumes that the hearer wants to convert the incoming stream of speech to an ordered output that will feed cumulative meaning composition. It also requires the parser to be able to track "height" in the functional sequence as an independent fact. This need not be problematic if all we assume the parser is sensitive to is whether the element on the stack "selects" the incoming element or not.

to offer us when it comes to the relative placement of specifiers and adverbials within the same hierarchical projection. Relevantly for us, Abels (2016) explicitly considers the question of whether adverb orders, head orders, and argument orders can form a joint sequence that is ordered by the algorithm and concludes that it only works if *each domain is sorted independently* – relative orderings within each class are predicted by the algorithm but contradictions arise if orderings across categories are included. This suggests that hierarchy is relevant to the embedding of complement projections inside the projection of the head, but that elements associated with the same projection must be internally differentiated in a different way. It also doesn't tell us how the grammar implements its own linearization rules formally and how they are cued for the learner and represented in the speaker's grammar. We will return to these questions for Scottish Gaelic in the final section of this chapter.

We have seen how a perspective of direct linearization uncovered a deep generalization about the mapping between hierarchy and word order in human languages, and how such a perspective can potentially lead to a simplification of the grammar in the form of the elimination of word order movements (phrasal roll up movements, in particular).

In the next subsection we will see how the perspective of direct linearization dissolves the problem of head movement.

4.1 Head movement

Since the onset of the Minimalist Program, head movement has been largely deprecated because it fails to "extend the target" in the statement of a syntactic derivation. There are many ways to rethink head movement so as to get around this problem.

Within direct linearization theories, there is a powerful early proposal for how to capture the descriptive effects of head movement, namely Brody's (2000) Mirror theory.

Observationally speaking, functional material which is hierarchically higher than the root is often linearized to the right of it (as a suffix), while the opposite linearization is observed outside of the word with higher functional material to the left. This tendency in many head initial languages was what led to head-movement and left-adjunction analyses of complex word formation (Baker 1985). It is quite common for languages to have distinct linearization principles for word internal bound morphemes than for the ordering of syntactically movable units. In fact, if we look at linearization as a core property of phonologization, we can

add the following criterion in (46) as a defining property of the phonological word.

(46) The phonological word as a linearization domain

The phonological word is the unit of the external linearization algorithm; it is a product of the internal linearization algorithm. Internal and external linearization are different mapping systems.

Note that the above statement does not attempt to relate the phonological word to atoms of the syntactic or semantic representation, but claims that the phenomenon of "word" in this sense emerges whenever a language has two distinct linearization cycles which feed each other.

In general, we can follow Brody (2000) in distinguishing the linearization statements that relate *one head to another directly subordinate one* from the linearization of "phrasal" elements. For the former, Brody (2000) gives the above direct linearization statement in (47) essentially mimicking in certain respects and replacing the rule of "head-movement".

(47) Word mirror (agglutinative head-head linearizations)

The syntactic relation 'X complement of Y' is identical to an inverse-order morphological relation 'X specifier of Y'.
(where the latter gives rise to the morphological structure [X [Y]] linearized from left to right). (after Brody 2000)

Brody's proposal is in some ways simply a different but equivalent implementation of head movement. There is however, one crucial innovation/flexibility opened up by this direct linearization approach. Under head movement, the pronounced word is spelled out at the landing site of movement, which is always the highest head position that is a part of the head word in question. The diacritic @ introduced by Brody (stated in (48) allows a headword formed by mirror (or spanning) to be in principle spelled out at *any* head position within the span, not just the highest position. This diacritic on the phrase structure can be used to regulate the height of Spell Out for a head within a language as a whole.

(48) Diacritic for height of Spell Out: (cf. Brody 2000)

In a tree, @ indicates the position where the morphological head word spells out.

Thus, in a tree such as Figure 2, the complex head formed from the combination of X and Y would always raise to the X position in a head-movement

implementation. But under the Brody proposal, the complex head so formed can be spelled out at Y as well, as a matter of parametric setting. This is indicated diacritically by the @ at Y, and indeed in cases of more extended sequences of heads forming head-words in this way, the @ can appear at the site of *any* of the heads in that sequence.

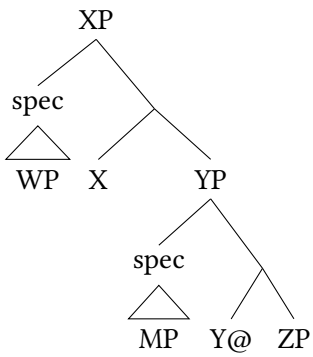


Figure 2: Height of Spell Out

Thus, if we think about the tree in Figure 2, the word order of the elements will follow the general mapping from c-command, with the caveat that the complex word [[Y] X] is spelled out as if it were in the Y position of the tree. The word order for Figure 2 with its diacritic would therefore be WP < MP < $[[Y]\ X]_{\omega}$ < ZP, whereas under head movement it would be WP < $[[Y]\ X]_{\omega}$ < MP < ZP.

In the spirit of Brody, I further assume that this principle of head-word formation applies also when the morphology in question is not strictly agglutinative, or even when the inflected word is completely synthetic. The point is that the head-word formation is a combination of the syntactic and semantic features of contiguous heads even when those are not spelled out by separable pieces of the morphology. The extension of head-word to non-agglutinative configurations is expressed in (49).

(49) Synthetic words, or adaptation to "spanning"

A morphological head span can also include spell-out of synthetic forms and not just agglutinative suffixation. (after Svenonius 2016, 2020)

With these tools in place, let us seek to represent the ordering of heads in Scottish Gaelic at the start of a clause. The relevant sentence is shown in (50) where I concentrate on the embedded clause 'that he did not speak with him.'.

(50) Thuirt Calum nach do bhruidhinn e ris.
say.PST Calum that.NEG PST speak.PST he with.3SM
'Calum said that he did not speak with him.'

In the embedded clause, C and Neg can be seen as being expressed by a synthetic head spanning C and Neg to give *nach*, T is separate here (*do*), as is the verb *bhruidhinn*. In the main clause, V-v-Asp-T is also a synthetic form *Thuirt*. A diacritic for linearization can be placed at Fin to ensure the high placement of the inflected verb in such cases. In the embedded clause, we need to allow separate Force and Neg heads, and allow a complex head-word to be formed from them that is separate from the complex head word involving tense. I will assume that the past tense particle *do* is in Fin, while the inflected verb itself is a head word consisting of V-v-asp-T and sits in T. In this system we see very clearly the ordering differences predicted for agglutinative morphology. If a tree like Figure 3 linearizes according to syntactic rules the functional head positions will line up in order with the top most functional particle leftmost. If on the other hand, a head-word is built a la Brody, then these heads will linearize within the word in the opposite order, and the actual height of the word itself will depend on the @ diacritic.

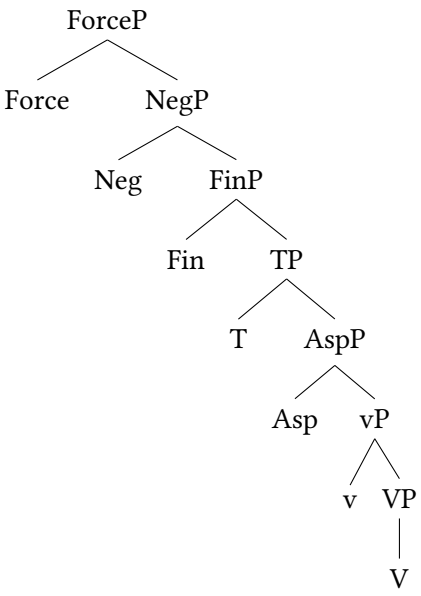


Figure 3: Tree with multiple CP level categories

We therefore need a separate diacritic to delineate stretches of heads that do form their own head-word linearization domain in this sense. In their discussion of spanning, Bye & Svenonius (2012) propose exactly such a diacritic to regulate the scope of Brody-type word formation in this sense (51).

(51) Scope of word formation (after Bye & Svenonius 2012)
A head marked with * indicates that the head so notated must form a Brody-type word with the closest head above it bearing an @ diacritic in the functional sequence.

For Scottish Gaelic therefore, we would mark the finite verb word as creating a span incorporating the V right up to T, linearized high at T. On the other hand, the Fin head is lexicalized separately, and the Neg and Force heads also form a head word together (Figure 4).

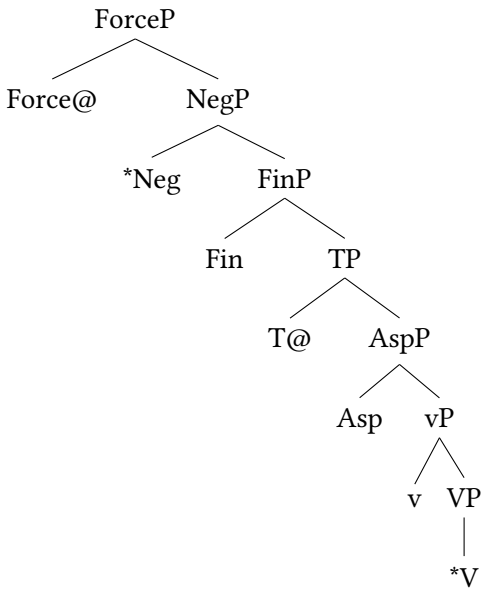


Figure 4: Linearization height for Scottish Gaelic

What the previous two subsections have shown us is that there are compelling direct linearization accounts that mimic the standard proposed syntactic movement accounts when it comes to the ordering of phrases and heads. However, these still fall short of delivering word order typology because they do not unify head, complement, specifier and adjunct ordering within a single system. The

stack sorting algorithms, as pointed out by Abels (2016), only work to deliver the right typological patterns if heads, arguments and adjuncts are ordered separately with respect to each other and not combined within a single sequence. For Brody's mirror, we are only dealing with heads by definition, and we make use of a diacritic to represent where complex heads interleave linearly with the extended complement structure. We don't in this system currently regulate the linearization of adjuncts for those projections. When it comes to Gaelic, the contiguity of the higher functional elements of the clause is dependent on the impossibility of adjunction to TP, for example.

5 Back to Scottish Gaelic word order

The current drawback in direct linearization accounts described so far is the failure to regulate the mutual linearization relationships between head, specifier, and adjunct within the same labeled projection. Indeed, it is not in the hierarchical ordering of functional projections per se that languages seem to differ (cartography has shown us remarkable similarity among languages with respect to functional sequence), but in the specific interleaving of heads and arguments, and the positioning of adverbials. It is here that languages make idiosyncratic, though systematic choices, and we need a way to describe these choices algorithmically in pursuit of a deeper understanding of the way in which hierarchical patterning in language interacts with serial externalization.

There is nothing principled in the way of pursuing a direct linearization approach to word order typology and variation; it is just that the project has not been widely embraced thus far. I hope that will soon change.

To summarize, we have seen that in Gaelic, the heads all conform to top down hierarchical order, but although C and T are obligatorily adjacent, they do not form a mirror order and so are not standardly analysed as involving head movement from T to C. There are no arguments or adjuncts that intervene within this stretch of heads. Adverbs are internally well behaved in that they line up in reverse scopal order when sentence final, but there is no statement in the grammar that explicitly prevents, for example, pre-adjunction to TP or to AspP in the absence of an overt head. With respect to arguments, subjects are always before objects except in copular clauses, something which echoes the typology of VSO languages more generally. Finally, light object pronouns seem to enjoy some right-linearization options, interleaved with adjunct positions. As I argued in Section 4, it seems that the head-parameter approach combined with a list of language specific obligatory movements does not even sufficiently describe these word order choices in Scottish Gaelic.

How then does a language decide to order its arguments and adjuncts with respect to a given head, and what are the constraints and cues to this mapping? In the last part of this chapter, I will speculatively pursue the idea that word order is a part of the syntax-phonology mapping and that phonologically motivated output constraints can be used to parametrize word order choices. The syntax-phonology interface is currently a theory of what the relationship is between the chunks, or constituents generated by the syntax, and the linearly sequenced chunks manipulated by the phonology. The basic question is therefore how syntactic phrases map onto prosodic units, given that the phonology seems to care about phonological words, phonological phrases (possibly recursive), and intonational phrases.[6]

While the syntax/phonology interface is a lively and productive subfield of linguistics, nobody seems to be actually trying to get word order to emerge from a language particular mapping, the phonologists seem to be leaving the actual order of words to the syntacticians. Since sound representations are inherently sequenced in human language, it makes sense that the first place that linear information would be present in language is after the mapping to phonology. Phonemes (or possibly phonological features) are sequenced into syllables which are then sequenced into phonological words. These latter units are now the units that are themselves linearized to give the neutral word order of the language. Bennett et al. (2016a) have argued that the placement of light object pronouns in Irish is triggered by a phonological requirement and is a part of the externalization mechanism. The position I will explore next, in line with direct linearization is that *all* word order arises from an externalization process and is therefore a property of the syntax-phonology mapping, not just this particular quirky word class.

(52) The phonological word as a linearization domain:
The phonological word is the input unit of the external linearization algorithm; it is a *product* of the internal linearization algorithm. Internal and external linearization are different mapping systems.

Given the language-specific construction of words, what sorts of algorithms take those units and the information of their hierarchical relationships and maps them to a linear representation?

[6]Disagreement exists with respect to whether the units of prosodic theory are related to the units of syntactic grouping by means of a *matching* algorithm between the units of syntax (XPs) and the units of prosody (phonological phrases ϕ) (Selkirk 2011), or whether the syntactic constituents just determine the *edges* of prosodic domains (Selkirk 1986 Nespor & Vogel 1982), creating many cases of mismatch.

In the next subsection, I will imagine the form of the mapping that would be needed to begin to make sense of some of the more problematic outstanding properties of Scottish Gaelic word order, including the placement of the subject immediately after the tensed verb, and the reduced availability of left adjoined positions for adverbials.

6 A Syn-Phon alternative to word order typology

From the discussion so far, we have seen that the key element in establishing a viable expression of a language's word order behaviours is to regulate the mutual ordering of heads, specifiers and adjunctions against the background of an assumed common hierarchy of phrase structural zones across languages (cf. Ramchand & Svenonius 2014).

To state the patterns, we must assume that the syntactic representation encodes hierarchical levels, and also encodes a difference between heads, specifiers and modifiers at each level. Given that this algorithm is a direct mapping between syntactic form and concrete serialization of phonologized elements of the structure, we must take note of whether a head of a particular phrase might be overtly filled or not, depending on independent language specific placements of the @ and * diacritics. For each phrase, there might be an argument in the specifier position or not, and there might be a modifier, or not.

Given that the mapping is specific to each language, and given also that it essentially orders phonologically overt elements with respect to each other, the output patterns of this mapping will have higher order phonological properties such as rhythm and stress. It is possible that some of the constraints on the output of the mapping can be expressed therefore as standard prosodic or phonological constraints since the output representation is essentially a linearized phonological representation. So, perhaps the space of those possible algorithms is even parametrized via phonological or rhythmic parameters on the output rather than overtly syntactic ones. Rhythmic cues to word order do not seem like an outlandish possibility. We know that child directed speech is particularly rhythmic and intonationally careful, and that children attend to the rhythm of the speech they are acquiring already in the womb (Nespor & Vogel 1986).

Quite generally, Scottish Gaelic phrases seem to consist of lone (relatively light) heads followed by substantial phrasal complements with very little left adjunction. We take this seriously in what follows.

Let us proceed top-down reading off the syntactic representation. The top of a complex syntactic representation will encode elements that belong to the top

tier of the hierarchy – the head, its related specifier, and any adjoined modifiers – as well as a phrase that contains all the embedded material, the complement. Following Chomsky (2000, 2001), I make use of the concept of the "edge" of a phrase to divide each constituent into two linearizable pieces: the "edge" and the complement, which I will call the "body" of the phrase. The edge is chosen here because it figures in many accounts of syntactic locality as an escape hatch for the impenetrability of phases, see the definition of the phase impenetrability condition (PIC) definition in (53). The edge seems to be the part of a phrase/phase whose Spell Out is delayed compared to that of the complement. In the definition below, the edge is claimed to consist of both the head and the material in the specifier. In my own implementation, I will include adjuncts to the phrase as being part of the edge.

(53) Phase Impenetrability Condition (PIC):

If Ψ is a phase with head H, then the complement of H (the "domain") is not accessible for operations involving a position outside Ψ. Only H and its specifiers (the "edge") are accessible for such operations.

In my algorithm, I separate each phrase into edge and body, where linearization of the former will be stipulated to precede the latter. While the edge is not itself a syntactic constituent, the "body" is, being the complement of the head. The edge however is well defined as the linear stretch bounded by the left boundary of the syntactic complement on its right, and the left boundary of the next constituent up. The first rule of Edge < Body linearization (Rule 1 below) must then be combined with a rule that orders elements *within* the edge if it happens to contain more than just a head (Rule 2). The body constituent is also complex of course, but then the whole cycle repeats recursively for body as the top node. The first rule of the cycle, Rule 1 is given in (54).

(54) Rule 1 (Recursive linearization rule):

For any syntactic phrase XP, with an overt edge and body portion, linearize the edge first followed by the body.
(Edge = Head and Phrase in specifier position, plus any adjoined phrases)
This is the part of the algorithm that overlaps with the LCA. It enforces a general effect of the functional sequence with higher material tending to precede lower material. However, it is recursive with respect to the body.

The algorithm does not yet exhaustively linearize all of the phonologically independent elements because edges themselves can be complex. In the maximal

case, they can consist of a head and an argument phrase in specifier position, as well as an adjoined modifier phrase.

Following the lessons from stack sorting, I will make the assumption that elements at the same height in this sense are not inherently ranked with respect to one another. The linearization algorithm is potentially sensitive to the phonological word vs. phonological phrase distinction, a phonological/prosodic distinction, but also potentially to any syntactic featural diacritic that is present on the syntactic representation it is interpreting. In particular, we should assume that the algorithm can see the distinction between a head and a phrase. My proposal here is that Scottish Gaelic is characterized by an essentially *iambic rhythmic choice* for complex edges. In other words, the head will always linearize before the phrase that is associated with it. Rule 2 is given in (55).

(55) Rule 2 (Complex edges):

If the edge is complex, i.e., consisting of both a head (a phonological word) and a phrase (a phonological phrase), then a step of linearization needs to occur within it as well.

A language can choose to solve that particular ordering statement either *iambically* or *trochaically*: Given that such a complex edge will consist of unequal prosodic constituents ω (phonological word) and ϕ (phonological phrase), either ω is ordered before ϕ (iamb), or vice versa (trochee).

In this case, Scottish Gaelic orders the head before the specifier. I assume this is what gives rise to the [Tense Subject] linearization unit and the [Asp Object] linearization unit. It is well known from prosodic work (see Elfner 2012) that the tensed verb and the subject form a close prosodic unit, and I speculate that this is correlated with the fact that they form part of the same subdomain for the purposes of linearization. For the purposes of illustration, I show how the linearization algorithm I propose would work on a toy example. The tree in Figure 5 shows the embedded clause linearized as *gun do leugh mi an leabhar*, where *gun* is the embedding complementizer, *do* is a past tense finiteness particle, *leugh* is the past form of the verb 'read', and *an leabhar* is the direct object 'the book'.

In the above tree, *gun* linearizes before FinP, by Rule 1, and *do* linearizes before TP by the application of Rule 1 on FinP. *Mi, leugh* together have to linearize before AspP, again by Rule 1, now applied to TP. The relative linearization of *mi* and *leugh* is resolved by the application of Rule 2, which says that the head must precede the specifier, to give *leugh-mi*.

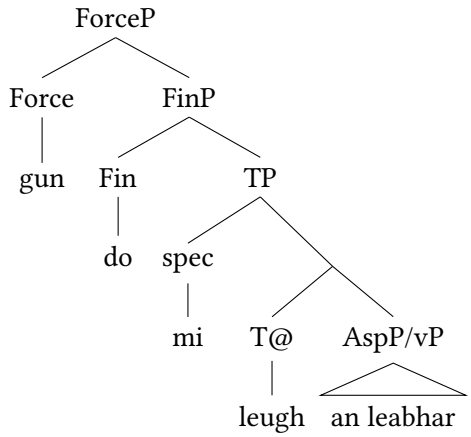


Figure 5: Linearization of an embedded clause

In the first obvious departure from standard accounts, the direct linearization theory does not need to assume that the tensed verb in Gaelic is sitting in a higher projection than the subject because of the VSO order. The claim here is that the subject is in the specifier of TP syntactically, as is assumed for most other tensed languages. VS order emerges because of the way that specifier head relation is *linearized* in Gaelic, not because of the verb moving past the subject.

Finally, we have seen multiple indications that modifiers are treated separately from arguments with respect to linearization, and so I will assume that specifiers (arguments) have a distinguished status here in having the option of being linearized within a complex edge according to Rule 2. To make the difference between specifiers and adjoined phrases, I will assume though that *non-selected* phrasal modifiers such as adverbials do not linearize in the same domain as the head as part of a core "complex edge". We will need a separate linearization rule for adjunct phrases.

There seems to be a fair amount of evidence that adjuncts are attached to particular projections or zones of the clause and can attach at various heights, but they often seem to linearize independently of their hierarchically matched argument DPs. When they are alone in the edge position they can linearize preceding the body that they dominate (as follows from Rule 1), but we need another rule in our arsenal to deal with adjuncts in syntactically heavy edges. This is given in (56).

(56) Rule 3 (Adjunct linearization):
Adjuncts can escape linearization at their point of hierarchical attachment.

> They can be eliminated from an edge and saved in a separate WORKING MEMORY BUFFER. Elements in the buffer are popped to the linearization string starting from the top of the stack, but only *after* the recursive body computation it belongs to.

In general, it seems reasonable to assume that the use of this working memory stack is parametrized across languages. Languages that do not tolerate heavy edges will place the offending adjunct in the BUFFER, when the edge contains other linearizable material. Languages on the other hand, might choose to tolerate heavy edges and have a linearization rule for them (for example, edges that contain both an argument and an adjunct but no overt head, could linearize trochaically as successive phrases as in English *Mary suddenly saw a bear.* or *Suddenly, Mary saw a bear.* Like stack sorting and roll up movements, buffering and popping at the end of a cycle of linearization, will produce an order of adjuncts *in reverse hierarchical order* at the end of the sentence. The notion of the buffer is thus a little bit like a special kind of stack, which preserves the order of the incoming input sequence (here driven by a top-down traveling through the syntactic tree) but then reverses it when translating into precedence. Given that the necessity of buffering is dependent on the head of the edge being filled, we can use this to model the occasional sentence medial adverbials we find in Scottish Gaelic. While a detailed analysis of all the cases is beyond the remit of this chapter, the intuition would be that in auxiliary constructions, because of the extra functional structure, the projection that the adverbial adjoins to has a phonologically null head position in the inflectional domain, which would allow the adjunct to linearize leftwards as part of a non-heavy edge. Alternatively, the choice between linearizing as part of the edge or undergoing buffering could be part of a systematic optionality for Scottish Gaelic adjuncts, with the latter being "preferred" under conditions that clearly require closer investigation. The point here is that a difference between the simple verb construction and the auxiliary verb construction with respect to adjunction positioning was ruled out in principle on a theory that simply reads off order from the abstract, non-phonologized phrase structure.

If we consider the possibility of adding an adverbial *gu slaodach* 'slowly', to the TP, Rule 3 will fail to linearize it before the TP material *an leabhar*, but will hold it in buffer until after that TP has been spelled out, and then order *gu slaodach* after that. If both a VP adverbial and a TP adverbial are present, then the VP adverbial will be popped out first, followed by the TP adverbial since the body linearization for VP completes before the full TP body linearization computation completes.

6.1 Characterising Scottish Gaelic word order

Scottish Gaelic in general imposes an *iambic* rhythmic pattern on its precedes relation for the phrasal algorithm. So edge precedes body, and within the edge itself, the word element (the filled head if there is one), will precede the selected phrasal argument in specifier position.

Adjuncts give rise to ”heavy edge” with filled heads, and are placed in the buffer, and linearized after the end of that recursive cycle. The recursive nature of the linearization algorithm and the different syntactic height of adverbials, forces adjuncts therefore to end up in reverse hierarchical order.

Although the phrasal prosodic algorithm for Scottish Gaelic is claimed here to be “iambic”, it is worth pointing out that at the prosodic word level, Gaelic is generally considered to be trochaic (Morrison 2019). This is not a contradiction because the domain of this parameter setting is phrasal linearization, not word internal linearization which is a completely separate cycle which leads to the definition of “word”’ in the first place as the elements that are then ordered by the later cycle. Scottish Gaelic is trochaic within the phonological word domain and iambic (if the word is even appropriate at these higher levels of abstraction) at the level of the phrasal linearization algorithm. It is possible in fact that the two domains will align oppositely with respect to this factor, as with ordering, and “mirror” (Brody 2000).

6.2 VOS as a sub-pattern

Given the algorithmic set up so far, we can make the following speculation: VOS is what happens in iambic languages when the Subject is linearized like adjuncts, via the buffer and so ends up last. The difference between linearizing within the edge grouping and buffering may be freer than I have assumed, and dependent rather on something syntactic like, for example, whether an Agree relation between the head and phrase has been established or not, or on some aspect of the intonation. The typological fact is that most VSO languages do seem to have VOS as a word order option. While this does not seem to be true on the face of it for Scottish Gaelic, it might well be that what we are seeing in the copular construction is a grammaticalized vestige of precisely this broader typological possibility.

One way of thinking about it, would be to say that the copula construction is different precisely in that the copula itself is base generated high and does not have an Agree dependency with the subject. This means that it will have to be buffered, just like an adjunct giving subject-last ordering for precisely this class of constructions.

(57) Is cìobar Calum.
Cop shepherd Calum
'Calum is a shepherd.'

If such a story could be made to work, it would also give possessor final order for possessed DPs, under the assumption that in Scottish Gaelic no formal Agree relation is established between the lexical noun and the possessor.

(58) car dearg a' bhalaich
car red the boy.GEN
'The boy's red car'

This strategy also allows for interleaved adverbs in the aspectual construction in Scottish Gaelic, where the verb and object similarly stay low, plausibly leaving the adverb alone in the edge of one of the intermediate functional projections.

6.3 The light pronoun problem

A crucial component of the Bennett et al. (2016b) analysis is the idea that object pronouns are too light to form their own phonological word. The present theory would make use of this idea as well, and assert that it makes the pronoun ineligible for edge linearization. It would therefore have to be thrown into the buffer, where it would linearize interleaved with adverbials, as observed.

6.4 What other linearization patterns are possible?

The rhythmic theory, together with the edge body partition give rise to a number of different word order setting possibilities. I explore some of these potential settings here.

For example, allowing trochaic ordering on the edge linearization rule, might systematically give rise to SVO. A relaxed tolerance for HEAVY EDGES might allow optionality on whether adverbials linearize via bufferization or not, but this would need more detailed examination of the patterns.

The general iambic rhythmic setting of Scottish Gaelic for edges is congruent with the properties of Rule 1, which orders head before complement in the simple case. If Rule 1 is specific to Gaelic, rather than being universal, it could be that head final languages would emerge from the opposite choice of linearization in Rule 1 making it the functional equivalent of the head parameter. However, Rule 1 is not a head ordering rule per se, since it orders the whole edge with respect to the body and not just the head. A reverse setting for Rule 1 would also give

languages that order the arguments after the heads that introduce them, in addition to having the heads appear in reverse hierarchical order. The overwhelmingly most common pattern for head final languages is for the arguments and adjuncts to precede all the verbal heads stacked up at the end, as if they resided in "leftward" specifiers. Another way to get head final languages from this parameter space would be to disallow prosodically unequal daughters in complex edges. This would force *heads* into the buffer making them linearize at the end of the externalization algorithm and stack up at the end in reverse hierarchical order. Then the edge would either be a single phrase, *or*, if it were complex, would just consist of two phrases (argument, and adjunct, say) which would be perfectly linearizable as equal daughters. Thus, Rule 1 is really the moral equivalent of the LCA, building in an asymmetric preference for height to correspond to leftness, while head finality is a parameter setting that affects the rhythm of the edge material. The iambic setting orders elements of unequal weight, the other setting prohibits unequal daughters and imposes an extrinsic order on things of equal weight (or maybe even allows both orders). The attractive thing about the approach being pursued here, to my mind, is the way that it utilizes factors that we already know the prosodic or rhythmic systems of language to be sensitive to. Bans on units that are too heavy, starting with things that are too light, unequal daughters etc. seem like reasonable constraints on a system regulating order via rhythm.

It is unclear from this kind of rhythmic set of parameters what the relationship is to the stack sorting algorithms that might underpin the mapping to hierarchy. Stack sorting gave us a handle on the general principles for reconstituting hierarchy from a range of different linearization choices. The rhythmic parameters explored here regulate the specifiers, heads, and adjuncts with respect to one another (although the invocation of the buffer is probably the equivalent of having a separate stack sorting mechanism for adjuncts). Also, this investigation has barely scratched the surface in terms of accounting for word order patterns cross-linguistically, or even completely for Scottish Gaelic itself, something beyond the scope of this short chapter. I hope instead to have achieved two modest goals. The first is to highlight the fact that the standard accounts currently fall short on accounting for some very general properties of Scottish Gaelic word order, including the positioning of adverbials (and the relative lack of left adjunction as a preferred option). The second goal was to open up the potential space of hypothesis formation and theory building to include theories of direct linearization. Direct linearization liberates the syntax from unmotivated movements and unwanted assumptions. It also allows for more phonological and prosodic considerations to make themselves felt if what we are dealing with in

externalization really is a kind of mapping, or interface, between syn/sem hierarchical representations and sound-based sequencing.

Acknowledgments

This chapter lived for a long time in basement of my brain before being finally written up. I am very grateful to Andrew Carnie for inviting me to present in his series at the University of Arizona, and thus giving me a chance to put my thoughts about word order typology into words. I am also grateful to various people over the years for having conversations with me about word order and syntactic movements. I would particularly like to single out David Adger, Klaus Abels (who introduced me to the work of David Medeiros), Peter Svenonius and Björn Lundquist. I would also like to thank an anonymous reviewer and Andrew Carnie for comments on an earlier draft, which greatly improved the chapter. They are not responsible for any remaining inaccuracies and infelicities. A big thank you finally to Björn Lundquist and Mike Hammond for help with LaTeX compatibility issues in the final stages.

References

Abels, Klaus. 2016. Neutral word order and stack sorting. ms. University College London.

Abels, Klaus & Ad Neeleman. 2009. Universal 20 without the LCA. In José Brucart, Anna Gavarró & Jaume Solá (eds.), *Merging features*, 60–80. Oxford: Oxford University Press.

Abels, Klaus & Ad Neeleman. 2012. Linear asymmetries and the LCA. *Syntax* 15(1). 25–74. DOI: 10.1111/j.1467-9612.2011.00163.x.

Adger, David. 1994. *Functional heads and interpretation*. University of Edinburgh. (Doctoral dissertation).

Adger, David. 2007. Pronouns postpose at PF. *Linguistic Inquiry* 38. 343–349. DOI: 10.1162/ling.2007.38.2.343.

Adger, David. 2010. Syntax. In Moray Watson & Michelle McLeod (eds.), *The Edinburgh companion to the Gaelic language*. Edinburgh: Edinburgh University Press.

Adger, David & Gillian Ramchand. 2003. Predication and equation. *Linguistic Inquiry* 34. 325–359. DOI: 10.21248/zaspil.26.2001.136.

Alexiadou, Artemis. 1997. *Adverb placement: A case study in antisymmetric syntax*. Amsterdam: John Benjamins.

Baker, Mark. 1985. The mirror principle and morphosyntactic explanation. *Linguistic Inquiry* 16(3). 373–415.

Bellert, Irena. 1977. On semantic and distributional properties of sentential adverbs. *Linguistic Inquiry* 8. 337–351.

Bennett, Ryan, Emily Elfner & James McCloskey. 2016a. Lightest to the right: An apparently anomalous displacement in Irish. *Linguistic Inquiry* 47(2). 169–234. DOI: 10.1162/LING_a_00209.

Bennett, Ryan, Emily Elfner & James McCloskey. 2016b. Lightest to the right: An apparently anomalous displacement in Irish. *Linguistic Inquiry* 47(2). 169–234. DOI: 10.1162/ling_a_00209.

Berwick, Robert C. & Noam Chomsky. 2011. The biolinguistic program: The current state of its development. In Cedric Boeckx & Anna Maria Di Sciullo (eds.), *The biolinguistic enterprise: New perspectives on the evolution and nature of the human language faculty* (Oxford Studies in Biolinguistics), 19–41. Oxford: Oxford University Press.

Biberauer, Theresa, Anders Holmberg & Ian Roberts. 2007. Disharmonic word-order systems and the final-over-final-constraint (fofc). In Antonietta Bisetto & Francesco Barbieri (eds.), *Proceedings of the XXXIII incontro di grammatica generativa*, 86–105.

Brody, Michael. 2000. Mirror theory: Syntactic representation in perfect syntax. *Linguistic Inquiry* 31(1). 29–56. DOI: 10.1162/002438900554280.

Bye, Patrik & Peter Svenonius. 2012. Non-concatenative morphology as an epiphenomenon. In Jochen Trommer (ed.), *The morphology and phonology of exponence: The state of the art*, 427–495. Oxford: Oxford University Press.

Carnie, Andrew. 1995. *Non-verbal predication and head-movement.* The Massachusetts Institute of Technology. (Doctoral dissertation).

Carnie, Andrew & Eithne Guilfoyle. 2000. Introduction. In Andrew Carnie & Eithne Guilfoyle (eds.), *The syntax of verb initial languages*, 3–12. Oxford University Press. DOI: 10.1093/oso/9780195132229.003.0001.

Carnie, Andrew, Heidi Harley & Elizabeth Pyatt. 2000. VSO order as raising out of IP? Some evidence from Old Irish. In Andrew Carnie & Eithne Guilfoyle (eds.), *The syntax of verb initial languages*, 39–60. Oxford: Oxford University Press. DOI: 10.1093/oso/9780195132229.003.0003.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In Roger Martin, David Michaels, Juan Uriagereka & S.J. Keyser (eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, 89–155. MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: A life in language*, 1–52. Cambridge: MIT Press. DOI: 10.7551/mitpress/4056.003.0004.

Chung, Sandra. 1983. The ECP and government in Chamorro. *Natural Language and Linguistic Theory* 1. 209–244. DOI: 10.1007/BF00154831.

Chung, Sandra. 1990. VP's and verb movement in Chamorro. *Natural Language and Linguistic Theory* 8(4). 559–620. DOI: 10.1007/bf00133693.

Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. New York: Oxford University Press.

Cinque, Guglielmo. 2004. Deriving Greenberg's Universal 20 and its exceptions. ms. University of Venice.

Cinque, Guglielmo. 2017. A microparametric approach to the head-initial/head-final parameter. *Linguistic Analysis*. 309–366.

Clemens, Lauren Eby & Maria Polinsky. 2013. Verb-initial word orders (primarily in Austronesian and Mayan languages). ms., Harvard University.

Doherty, Cathal. 1996. Clausal structure and the Modern Irish copula. *Natural Language and Linguistic Theory* 14(1). 1–46. DOI: 10.1007/bf00133402.

Doron, Edit. 2000. VSO and left-conjunct agreement: Biblical Hebrew vs. Modern Hebrew. In Andrew Carnie & Eithne Guilfoyle (eds.), *The syntax of verb initial languages*, 75–95. Oxford: Oxford University Press. DOI: 10.1093/oso/9780195132229.003.0005.

Elfner, Emily. 2012. *Syntax-prosody interactions in irish*. Santa Cruz, CA: UC Santa Cruz. (Doctoral dissertation).

Elfner, Emily. 2015. Recursion in prosodic phrasing: Evidence from Connemara Irish. *Natural Langauge and Linguistic Theory* 33. 1169–1208. DOI: 10.1007/s11049-014-9281-5.

Emonds, Joseph E. 1980. Word order and generative grammar. *Journal of Linguistic Research* 1. 33–54.

Freeze, Ray & Carol Georgopolous. 2000. Locus operandi. In Andrew Carnie & Eithne Guilfoyle (eds.), *The syntax of verb initial languages*, 164–183. Oxford: Oxford University Press. DOI: 10.1093/oso/9780195132229.003.0009.

Greenberg, Joseph H. 1966. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of language. Report of a conference held at Dobbs Ferry, NY, april 13-15, 1961*, 2nd edn., 73–113. Cambridge: MIT Press.

Guilfoyle, Eithne. 1990. *Functional categories and phrase structure parameters*. McGill University. (Doctoral dissertation).

Haider, Hubert. 2000. OV is more basic than VO. In Peter Svenonius (ed.), *The derivation of VO and OV*, 45–67. Amsterdam: John Benjamins.

Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar* (Current Studies in Linguistics 2). Cambridge: MIT Press.

Kayne, Richard S. 1994. *The antisymmetry of syntax*. Cambridge: MIT Press.

Knuth, Donald. 1968. *The art of computer programming*. Reading: Addison-Wesley.

Koopman, Hilda & Anna Szabolcsi. 2000. *Verbal complexes*. Cambridge: MIT Press.

McCloskey, James. 1979. *Transformational grammar and model theoretic semantics*. Dordrecht: Kluwer.

McCloskey, James. 1990. Resumptive pronouns, A-bar binding and levels of representation in Irish. In Randal Hendrick (ed.), *The syntax of the modern Celtic languages: Syntax and semantics 23*. New York: Academic Press.

McCloskey, James. 1991. Clause structure, ellipsis and proper government in Irish. *Lingua* 85. 259–302. DOI: 10.1016/0024-3841(91)90023-x.

McCloskey, James. 1996. On the scope of verb movement in Irish. *Natural Language and Linguistic Theory* 14(1). 47–104.

McCloskey, James. 2005. A note on predicates and heads in Irish clausal syntax. In Andrew Carnie, Heidi Harley & Sheila Ann Dooley (eds.), *Verb first: On the ordering of verb initial languages*, 155–174. Benjamins. DOI: 10.1075/la.73.12mcc.

Medeiros, David. 2018. ULTRA: Universal grammar as a universal parser. *Frontiers in Psychology: Language Sciences* 9. 151–183. DOI: 10.3389/fpsyg.2018.00155.

Medeiros, David. 2024. Stack sorting grammar. ms. University of Arizona.

Morrison, Donald A. 2019. Metrical structure in Scottish Gaelic: Tonal accent, glottalisation and overlength. *Phonology* 36(3). 391–432. DOI: 10.1017/S0952675719000204.

Myhill, John. 1985. Pragmatic and categorial correlates of VS word order. *Lingua* 66. 177–200. DOI: 10.1016/s0024-3841(85)90314-6.

Nespor, Marina & Irene Vogel. 1982. Prosodic domains of external Sandhi rules. In Harry van der Hulst & Norval Smith (eds.), *The structure of phonological representations*, 225–255. Dordrecht: Foris.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology* (Studies in Generative Grammar). Dordrecht, Holland; Riverton, N.J.: Foris.

Noonan, Maire. 1994. VP internal and VP external AGROP. In Raul Aronovich, William Byrne, Susanne Preuss & Martha Senturia (eds.), *The proceedings of the thirteenth West Coast conference on formal linguistics*, 318–333. San Diego: University of California, San Diego.

Ramchand, Gillian. 1997. *Aspect and predication: The semantics of argument structure*. Oxford: Oxford University Press.

Ramchand, Gillian & Peter Svenonius. 2014. Deriving the functional hierarchy. *Language Sciences* 46. 152–174. DOI: 10.1016/j.langsci.2014.06.013.

Roberts, Ian. 1996. *Comparative syntax*. London: Edward Arnold.

Selkirk, Elisabeth. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3. 371–405.

Selkirk, Elisabeth. 2011. The syntax phonology interface. In John A. Goldsmith, Jason Riggle & Alan Yu (eds.), *The handbook of phonological theory*, 435–484. Oxford: Wiley Blackwell. DOI: 10.1002/9781444343069.ch14.

Shlonsky, Ur. 1997. *Clause structure and word order in Hebrew and Arabic: An essay in comparative Semitic syntax*. Oxford: Oxford University Press.

Sportiche, Dominique. 1988. A theory of floating quantifiers and its corollaries for constituent structure. *Linguistic Inquiry* 19(3). 425–449.

Svenonius, Peter. 2016. Spans and words. In Dan Siddiqi & Heidi Harley (eds.), *Morphological metatheory*, 201–222. Amsterdam: John Benjamins.

Svenonius, Peter. 2020. A span is a thing: A span-based theory of words. *NELS 50: Proceedings of the Fiftieth Annual Meeting of the North East Linguistic Society* 3. 193–206.

Chapter 4

Reassessing Oehrle effects: Evidence from Scottish Gaelic

Gary Thoms
New York University

It has often been argued that the double object construction and prepositional dative construction differ semantically – in particular, that the double object construction has a possessive core which is not shared by the prepositional dative – on the basis of various cases where the dative alternation breaks down in languages such as English. These "Oehrle effects" have been argued to support a two-base analysis of the dative alternation (e.g., Harley 2002) over derivational accounts (e.g., Larson 1988). This article counters this argument on the basis of data from Scottish Gaelic, which expresses ditransitives exclusively in terms of prepositional datives. It is shown that Oehrle effects do not obtain in this language: various ditransitive types which resist the prepositional dative in alternating languages do not show any such resistance to the frame in Scottish Gaelic, a fact that tells us that the problem behind Oehrle effects cannot be one of semantic incompatibility. I provide an alternative syntactic account of Oehrle effects in English and their absence in Scottish Gaelic, arguing that they can be explained in terms of the height of certain subject arguments and their interaction with object shift operations, which vary from language to language.

1 Introduction

In Scottish Gaelic and Irish, ditransitives are expressed exclusively in the form of prepositional dative constructions (PDCs), as in (1–2), where the indirect object (IO; the goal/recipient) is marked with a preposition and preceded by an unmarked direct object (DO; theme) in neutral word order.

Gary Thoms. 2025. Reassessing Oehrle effects: Evidence from Scottish Gaelic. In Andrew Carnie, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond & Luis Irizarry (eds.), *Foundational approaches to Celtic linguistics*, 141–181. Berlin: Language Science Press. DOI: 10.5281/zenodo.15654863

(1) Thug Màiri leabhar do dh'Iain.
give.PST Mairi book to Iain
'Mairi gave a book to Iain.' (SG)

(2) Thug Máire leabhar do Sheán.
give.PST Maire book to Sean
'Maire gave a book to Sean.' (Irish)

Most theoretical analyses of PDCs in the literature have centered on the question of whether they are related to double object constructions (DOCs), in particular in languages such as English that show an alternation between the two, known as the dative alternation. In one family of analyses, the two constructions are taken to be derivationally related, and one of the two structures forms the base for construction of the other; we can call this the *derivational theory* (Larson 1988; Snyder & Stromswold 1997; Ormazabal & Romero 2012, 2023; Collins 2024; cf. also Rappaport Hovav & Levin 2008). In another family of analyses, which Harley & Miyagawa (2017) identify as being representative of a "relatively broad consensus", the two constructions are distinct argument structures altogether with different base positions for the arguments; call this the *two-base theory* (Green 1974; Oehrle 1976; Marantz 1993; Pesetsky 1995; Pylkkänen 2002, 2008; Harley 2002; Harley & Jung 2015; Bruening 2001, 2010a, 2010b, 2018). The debate between these two theories speaks to a number of major theoretical issues, such as the status of Baker's Uniformity of Theta Assignment Hypothesis (UTAH): in essence, UTAH predicts that some version of the single-base theory must be right; it would, in its strongest form, rule out the two-base theory altogether. The debate has largely centered on alternating languages such as English, while relatively little attention has been paid to PDC-only languages such as Scottish Gaelic and Irish.

One of the main sources of debate is the proper treatment of cases where the dative alternation breaks down, as in (3–5), or where the two frames seem to give rise to different interpretations, as in (6). Many such cases were first identified in English by Oehrle (1976; but see also Green 1974), and they have often been described as "Oehrle effects".

(3) a. The editor sent the article to Philadelphia.
b. *The editor sent Philadelphia the article.

(4) a. The war years gave Mailer a book.
b. *The war years gave a book to Mailer.

(5) a. The lighting here gives me a headache.
b. *The lighting here gives a headache to me. (Bruening 2010a: 288)

(6) a. John gave Mary a child. *can mean* 'John got Mary pregnant'
b. John gave a child to Mary. *can't mean* 'John got Mary pregnant'
(Harley 2002: 42)

A common analysis of these pairs is that they show that the DOC and PDC have distinct meanings: DOCs have a CAUSED POSSESSION interpretation, whereby the indirect object (IO) becomes the possessor of the direct object (DO), while PDCs don't involve caused possession but rather some event involving an agent, a theme, and a goal that is introduced by a preposition. Ditransitives of the kind in (3–6) allow us to see this difference, it's claimed, because one or the other of their arguments is semantically incompatible with the relevant thematic roles; for example, the IO in (6) must be a true possessor for the intended (inalienable) interpretation.[1] This account makes a clear prediction, mentioned by Harley (2002: 66, fn. 12): in languages that don't have access to the DOC at all, such as the Celtic languages, we should find that the Oehrle effect cases – those ditransitive types that are specific to DOCs, such as (4–6) above – are simply inexpressible with the language's ditransitives because the PDC frame is semantically incompatible with such caused possession meanings.

In this chapter, I show that this prediction is not borne out: PDCs in non-alternating languages such as Scottish Gaelic don't show any of the supposedly semantic restrictions that English PDCs show. It can be concluded that the English contrasts should not be explained in terms of semantic distinctions between the frames but rather in terms of some other, most likely syntactic, ingredients of the ditransitive frames. With the semantic explanation put to one side, I develop an alternative syntactic analysis of Oehrle effects, attributing the relevant effects to the interaction of object shift and low non-agentive subjects. The differences between English and Scottish Gaelic that are documented with respect to Oehrle effects are then tied to differences in the distribution of object positions in the languages. The resulting picture is one where the apparently semantic constraints on PDCs are given a thoroughly syntactic reappraisal.

[1] This reading seems somewhat archaic for many speakers, but a similar effect can be demonstrated with *John and Mary gave their parents a beautiful grandchild*, which is less archaic and also fails as a PDC on the right interpretation (**John and Mary gave a beautiful grandchild to their parents*).

2 On the two-base theory

The leading idea in Harley (2002) is that ditransitives involve a causative light verb embedding a small clause-like structure which is headed by a preposition, and the preposition introduces one argument in its complement domain and another in its specifier. The two argument structure frames involve the use of different abstract prepositions which have distinct but largely overlapping meaning contributions, and they are taken to be related to the structures involved in different kinds of possessive predication crosslinguistically. For DOCs, the preposition is identified with the P that Kayne (1993) took to be at the core of HAVE-possessives – hence it is called P_{HAVE} – and it takes a DP as its complement. This is in Figure 1a. For PDCs, the preposition has a locative semantics – hence it is called P_{LOC} – and it takes a PP as its complement. This preposition, Harley claims, is akin to the one that is involved in negotiating predicative possession relations in BE-possessive languages, such as Irish and Scottish Gaelic. This is in Figure 1b.

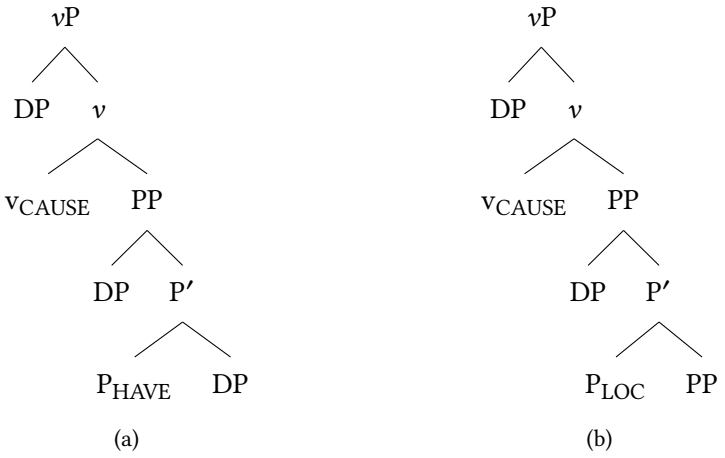


Figure 1: Harley's structures for DOCs (a) versus PDCs (b)

Harley's account captures contrasts such as (3–5) quite straightforwardly. (3) follows because "Philadelphia" can be a location or goal but not a possessor (cf. *#Philadelphia has the letter*), whereas (4 and 5) hold because the predications involve some sort of caused possession (e.g. 5a is paraphrasable as "the lighting caused me to have a headache"); however, the IOs are not locations or goals in any real sense.

Harley's analysis generalizes to other cases of failed PDCs where the IO is not plausibly a location or goal, whereas it could be a possessor in the associated

DOC, as is the case with *cost* and *spare* (7–8), both of which have negative caused possession interpretations of some sort ("cause to not have").

(7) a. That decision cost John a job.
b. *That decision cost a job to John.

(8) a. They spared me that ordeal.
b. *They spared that ordeal to me.

It may also capture some effects noted by Green (1974: 157) regarding the interpretation of the ditransitive use of *teach* as in (9), where it is claimed that while the DOC in (9a) has the implication that the students ended up learning ('having') some French, the PDC in (9b) lacks this implication. Although the effect in (9) is weak or nonexistent for some speakers or in certain configurations,[2] the caused possession interpretation is arguably more strongly required with cases such as (10a) with non-agentive causer subjects, where the corresponding PDCs (10b) are wholly unacceptable, as noted by Oehrle (1976).

(9) a. John taught the students French.
b. John taught French to the students.

(10) a. Lipson's textbook taught me Russian.
b. *Lipson's textbook taught Russian to me. (Pesetsky 1995: 193)

These effects should follow if the PDC lacks a caused possession interpretation, and instead its PP has a goal interpretation.

Another set of contrasts that Harley attributes to the two-base theory is the fact that many ditransitive idioms fail to alternate between one frame or another (see also Richards 2001). (11–12) demonstrate two well-known examples.

(11) a. Oscar will give John the boot.
b. *Oscar will give the boot to John.

(12) a. Lasorda sent his starting pitcher to the showers.
b. *Lasorda sent the showers his starting pitcher.

[2]See Rappaport Hovav & Levin (2008: 144–150) for discussion of this specific example, and also the general question of whether DOCs and PDCs really do differ systematically with respect to successful transfer inferences.

Harley argues that such facts can be captured if the idioms involved correspond to stored substructures of the distinct prepositional small clauses associated with the two frames. So, *give X the boot* involves a P_{HAVE}' idiom (comprising the P_{HAVE} head and its complement DP *the boot*), and so (11b) is impossible because the PDC does not have P_{HAVE}' as its substructure. Something similar is said to obtain with (12a), where the idiom involved is a P_{LOC} idiom which cannot be a subpart of a DOC, hence the impossibility of (12b).

Subsequent incarnations of the two-base approach, such as Bruening (2010a, 2010b) and Harley & Jung (2015), have maintained the same basic idea: the DOC involves a caused possession semantics whereas the PDC does not. Bruening argues against the "symmetry" of Harley's small clause analysis, and in favour of an analysis where DOCs are structurally more complex – with an ApplP projection above VP encoding possessive semantics, whereas the PDC lacks this and, thus, the possessive interpretation. Most of the issues discussed in what follows present the same sort of challenges for this whole set of analyses; therefore, I focus on Harley's original proposal with some additional remarks clarifying differences with others.

3 Celtic and Oehrle effects

3.1 Celtic and the distribution of ditransitive frames

The Celtic languages are all VSO languages whose ditransitives involve V-S-DO-IO order by default. In Scottish Gaelic, IOs can be introduced by various prepositions, but the most widely used ones are *do* (plus lenition) 'to' and *gu* 'to(wards)', the latter typically conveying a clearer sense of directed motion; compare (13) with (14). The light verb *thoir* combines with either, giving rise to different translation equivalents in each case. (Henceforth, all non-English examples are Scottish Gaelic unless otherwise indicated.)

(13) Thug iad leabhraichean do dh'Iain.
give.PST they books to Iain
'They gave books to Iain.'

(14) Thug iad leabhraichean gu Iain.
give.PST they books to Iain
'They brought books to Iain.'

As noted, *gu* 'to(wards)' is typically used for verbs with a directed motion component, similar to *cuir* in its 'send' use (see 15–16), but a distinct preposition *a* is used

with locational goals. This is most likely a distinction akin to the dative/allative distinction, the relevance of which for the analysis of the *send* contrasts is insightfully discussed in Rappaport Hovav & Levin (2008).

(15) Chuir iad leabhraichean gu Iain.
send.PST they books to Iain
'They sent books to Iain.'

(16) Chuir iad leabhraichean a Ghlaschu.
send.PST they books to Glasgow
'They sent books to Glasgow.'

Prepositions in these languages inflect for features of their complements, and pronominal complements are covert due to the well-known complementarity between head marking and pronominal expression we see in these languages. In what follows I simply gloss the P-pronoun amalgamations in the format 'at.me' and 'with-him' for ease of reading, with no concrete commitment on the specific analysis (on which see e.g., McCloskey & Hale 1984).

Harley (2002) brought Celtic into the discussion in the ditransitive literature by way of a typological argument in favour of her theory. Having proposed P_{HAVE} as the crucial ingredient of DOCs, Harley identified the prediction that languages which lack a verbal HAVE predicate will also lack DOCs, since both are based on P_{HAVE}. She claimed that Irish and Scottish Gaelic bore this out, since they lack verbal HAVE and also lack DOCs, as the examples in (17) show.

(17) a. *Thug Máire S(h)eán leabhar.
give.PST Maire Sean book
intended: 'Maire gave Sean a book' (Irish, Jung et al. 2012)

b. *Thug Mairi (dh')Iain leabhar.
give.PST Mairi Iain book
intended: 'Mairi gave Iain a book.'

The Celtic languages express predicative possession with BE-possessives that mark the possessor with an adposition. In Scottish Gaelic, the adposition is *aig* 'at', and the PP shows up after the possessee (18).

(18) Tha leabhar aig Iain.
be.PRS book at Iain
'Iain has a book.'

Harley (2002) claims that this constitutes an instance of the P_{LOC} small clause that is implicated in her analysis of the PDC.

However, there are some challenges for this analysis. One comes from the fact that the BE-possessive clauses in Celtic seem to be very much possessive and not locative in their semantics. For example, they can be used to express inalienable possession relations of various kinds, as well as abstract possession, as the examples in (19) demonstrate for Scottish Gaelic.[3]

(19) a. Tha piuthar agam.
be.PRS sister at.me
'I have a sister.' (somewhere in the world)

b. Tha sùilean gorma agam.
be.PRS eyes blue at.me
'I have blue eyes.'

c. Tha ceann goirt agam.
be.PRS head sore at.me
'I have a headache.'

In fact, the language even uses the same construction for having knowledge of a language. If the PP argument of a P_{LOC} structure is a goal, as Harley claims, then it's hard to see how (20) could be well-formed whereas (10b) with *teach* in the PDC could be bad. Recall, the latter was ruled out on the basis of the PP argument that represents the seat of imparted linguistic knowledge not making a good goal, yet by the same logic the PP in (20) would be the same sort of of goal.

(20) Tha Gàidhlig agam.
be.PRS Gaelic at.me
'I know Gaelic', 'I have Gaelic', lit. 'Gaelic is at me'

If the language's PDCs and predicative possession clauses have common grammatical ingredients, one would not really expect the PDCs to be restricted to locative interpretations.

[3]Part-whole relations with inanimates are expressed with similar constructions but a different preposition, *air* "on".

(i) Tha ceithir casan air a' bhòrd.
be.PRS four legs on the table
'The table has four legs.'

Still, the relationship between the BE-possessives and the PDCs must be at least a little abstract, because they involve different adpositions (*aig* 'a' with BE-possessives, *do* and others in PDCs). Therefore, it is possible that these independent differences account for the difference between predicative possession and Scottish Gaelic PDCs with respect to their possessive interpretations. However, we still need to ascertain whether Scottish Gaelic PDCs do have a possessive interpretation when looking beyond the simple examples discussed so far.

3.2 Caused possession

Harley (2002: 67, fn 12) noted the specific prediction that the Celtic languages should not allow caused possession constructions of the type associated with Oehrle effects, saying "[o]ne wonders if in Irish idiomatic/metaphoric expressions like *The war years gave Mailer a book* are ill-formed; the current analysis seems to predict that they should be." Although this particular construction doesn't seem to arise in Scottish Gaelic,[4] similar examples with caused possession of abstract DOs are possible, such as (21) and (22).

(21) Thug a bhliadhnaichean sa chogadh sealladh ùr dha.
gave his years in.the war outlook new to.him
'His years in the war gave him a new outlook.' (cf. **His years in the war gave a new outlook to him*)

(22) Thug na thuirt thu beachd dhomh.
give.PST the.PL said you idea to.me
'What you said gave me an idea.' (cf. **what you said gave an idea to me*)

Intriguingly, the construction in (22) can also be rendered with the directional preposition *gu*, which typically translates as 'towards'. (*Thugam* is the 1SG-inflected form of *gu*.)

(23) Thug e beachd thugam
give.PST it idea to.me
'It gave me an idea.'

[4]This is paralleled to some degree by the facts from English; I have found that a reasonable number of speakers (including native-speaker linguists who are familiar with the issues) find these *give X (the idea for) a book* uses of *give* quite outside of their English. I have not investigated this variation in sufficient detail to rule out the possibility that there is some independent issue with these specific sentences. All of these speakers find a version of the DOC with *an idea for a book* as the DO completely acceptable.

There are also cases such as *give a headache*, which seem to involve caused possession with non-animate, non-agentive subjects (24).[5] Similar examples which involve giving people diseases or illnesses are also possible, (25).

(24) Thug an fhuaim ceann goirt dhomh.
give.PST the noise head sore to.me
'The noise gave me a headache.' (cf. **the noise gave a headache to me*)

(25) Thug Iain covid dhomh.
give.PST Iain covid to.me
'Iain gave me covid.' (cf. **Iain gave covid to me*)

It seems, then, that *thoir* can combine with *do* and sometimes even *gu* to give rise to unambiguous caused possession interpretations in Scottish Gaelic. This in turn suggests that PDC frames are not semantically incompatible with such interpretations.

We can also consider uses of *give* with kinship terms such as *daughter/son*, where the external argument is not an agent but rather some sort of causer, and what is caused is an inalienable possession relation. These strongly resist the PDC on the relevant interpretation in English; see (6) above, where only the DOC has the inalienable possession interpretation. This follows on Harley's analysis from the fact that these require a true (inalienable) possession interpretation, and the PDC doesn't allow for such an interpretation.[6] Yet these examples are unproblematic in the PDC in Scottish Gaelic with *thoir*.

(26) Thug i nighean dha.
give.PST she daughter to.him
'She gave him a daughter.' (**she gave a daughter to him*)

Again, the prediction of Harley's proposal is not borne out.

The effect of the "Lipson's textbook" examples can be tested with *teach/learn*-type verbs. The examples in (27) show that Scottish Gaelic allows the use of the PDC in such cases; *theasaig*, which translates more directly as 'taught', would work in these cases too.

(27) a. Dh'ionnsaich an cùrsa sin beagan Gàidhlig dhomh.
learn.PST the course that little Gaelic to.me
'That course taught me a little Gaelic.'

[5]Note that *sore head* is a common idiomatic expression for headache in Scottish English.

[6]Examples such as (26) are just as archaic for my Scottish Gaelic informant as they are for English speakers, as noted in footnote 1.

b. Dh'ionnsaich an leabhar sin còig faclan Gàidhlig dhomh.
learn.PST the book that five words Gaelic to.me
'That book taught me five Gaelic words.'

c. Dh'ionnsaich an t-òran sin co-fhaireachdainn dhomh.
learn.PST the song that empathy to.me
'That song taught me empathy.'

These were cases that strongly resisted the PDC frame in English, with causer subjects and IOs that do not make particularly clear goal arguments. What was somewhat unclear from these examples, and Oehrle effects more generally, is whether the problem with the PDC is the goal semantics of the IO or the non-agent status of the external argument. Given that non-agentive subjects are in principle compatible with goal arguments in the VP (e.g., *the wind moved the desk to the other side*), it seems more likely that the real issue with the English cases is the external argument's thematic status rather than the status of the IO. This doesn't obviously follow from the Harley account, in which the external argument is introduced in Spec,*v*P and all of the action takes place in the small clause complement of *v*.

3.3 Non-goal indirect objects

Looking beyond the core Oehrle effects cases, we can broaden our view to consider other ditransitives where the IO is not plausibly a goal argument. Such cases are usually not possible in PDCs in English, and Harley's account readily explains them. The question, then, is whether this holds in Celtic. One class of cases involves semelfactive nouns such as *pòg* 'kiss' (28a) and *breab* 'kick' (28b), which are quite productive in light verb constructions in Scottish Gaelic. The English PDCs are degraded (while the DOCs are fine), and this follows if the PP is necessarily a goal. However, it then leaves the acceptability of the Scottish Gaelic PDCs to be explained.

(28) a. Thug e pòg dhomh.
give.PST he kiss to.me
'He gave me a kiss.' (cf. **he gave a kiss to me*)

b. Thug i breab dha.
give.PST she kick to.him
'She gave him a kick.' (cf. **she gave a kick to him*)

Exactly how these constructions are to be dealt with under the P_{HAVE} account isn't clear; Harley (2002: 39) mentions English examples such as (28b), but doesn't state an analysis. If they are idiomatic, they would be analyzed as P_{HAVE} idioms (which we come back to shortly). If they are productive and compositional, then presumably P_{HAVE} would be displaying some of the flexibility that HAVE shows in examples such as *I'll have a shower/look*. I don't commit to a full analysis here, but merely note that the contrast between Scottish Gaelic and English with respect to the PDC variants is in need of some explanation, and it doesn't follow in an obvious way from Harley's account.

An additional set of cases to consider is verbs such as *cosg* 'cost' and *diùlt* 'refuse/ deny', which strongly resist a PDC frame in English, especially if the DO is an abstract possessum, as we saw in (7b) above. In such examples, the IO is not plausibly a goal, and yet the PDC is fine in Scottish Gaelic with such a verb, as shown in the examples in (29).

(29) a. Chosg a' cho-labhairt tri mìle not dha.
cost.PST the conference three thousand pound to.him
'The conference cost him three thousand pounds.' (Adger 1994)

b. Chosg e ùine dha.
cost.PST it time to.him
'It cost him time.' (cf. **It cost time to him*)

c. Dhiùlt e mo dhinneir dhomh.
refused he my dinner to.me
'He refused me my dinner.' cf. *(*He refused my dinner to me*)

Bruening (2010a: 558) sketches a semantics for *spare* in English (as in 8) that makes crucial use of the possession role encoded by Appl, and that rules out appearing with the PDC. Something similar ought to apply to *cost* and *refuse/deny*, in which case the Gaelic example would once more show the availability of a possessive interpretation with PDCs.

3.4 Ditransitive idioms and the *give/get* alternation

Scottish Gaelic has a number of ditransitive idioms which feature an idiomatic DO, an open (transparent) IO argument position and a preposition selecting the IO which is specific to the idiom. A handful are given in (30) with pronominal IOs, but they are just as acceptable with full DPs as complements to the prepositions.

(30) a. Thug i a' bhròg dha.
give.PST she the shoe to.him
'She fired him.', lit. 'She gave the shoe to him.'[7]

b. Thug i cothrom dha.
give.PST she chance to.him
'She gave him an opportunity.', lit. 'She gave a chance to him.'

c. Thug iad feannadh air.
give.PST they disemboweling on.him
'They gave him a tongue-lashing.', lit. 'They gave a disemboweling to him.'

d. Thug i ruith air.
give.PST she running on.him
'She gave him a telling off.', lit. 'She gave a running on him.'

e. Thug i a' char às.
give.PST she the twist out.of.him
'She tricked him.', lit. 'She took the twist out of him.'

In addition, *ceann goirt* 'sore head' from above can also have the idiomatic meaning 'problem', as in (31).

(31) Thug Iain ceann goirt dhomh.
give.PST Iain head sore to.me
'Iain gave me a problem (to deal with).'

These pose different kinds of problems for the different analyses of ditransitive idioms in Harley (2002) and Bruening (2010a).[8]

For Harley, the idioms here look more like what she describes as P_{HAVE} idioms, since they involve an idiomatic DO and a transparent non-idiomatic IO; indeed (30a) is very similar to *give X the boot*, so much so that we might speculate that they're related historically. The *do*-based idioms in (30a–30b) also share another thing in common with English P_{HAVE} idioms: they also occur with their idiomatic meaning with a verb translating as *get*, where the IO argument occurs as the subject and the idiom chunk occurs as the complement of the verb. The English examples in (32) and (33) demonstrate this effect; Richards (2001) explains it by

[7] Much like the English equivalent with *boot*, the idiomatic interpretation disappears if there is no definite article: *thug i bròg dha* can only have the literal meaning "she gave him a shoe".

[8] See also Larson (2017) for a broader critique of the argument for the two-base theory from ditransitive idioms.

positing a decompositional analysis of *get* where it contains the same P_{HAVE} that occurs in *give* (the two differing only with respect to the content of *v*).

(32) a. Oscar will give John the boot. (=11a)
b. His advisor really gave John a kick in the pants.
c. Susan gave Bill a piece of her mind.

(33) a. John got the boot.
b. John got a kick in the pants.
c. Bill got a piece of my mind.

The Scottish Gaelic verb *faigh* (past tense *fhuair*) translates quite directly as *get*, and the following examples in (34) and (35) show that the *thoir*-based idioms that occur with the preposition *do*[9] retain their idiomatic meanings in combination with *faigh*, with the arguments that occurred as prepositionally marked IOs occurring as the subject.

(34) a. Fhuair e a' bhròg.
get.PST he the shoe
'He got fired.'
b. Fhuair Iain cothrom (bho Mhàiri).
get.PST Iain chance from Mairi
'Iain got an opportunity (from Mairi).'
c. Fhuair Iain ceann goirt.
get.PST Iain head sore
'Iain got a sore head.' or 'Iain got a problem (to deal with).'

(35) Fhuair mi beachd (bho Mhàiri).
get.PST I idea from Mairi
'I got an idea (from Mary).' or 'I got an opinion (from Mary).'

[9]Interestingly, the non-*do*-based idioms do not retain their idiomatic meaning when inserted into the *faigh* frame.

(i) *Fhuair iad feannadh.
get.PST they disemboweling
intended: 'They got a tongue-lashing.'

I speculate that this is because the preposition that is implicated in their idiomatic interpretation is not retained, but I refrain from spelling out a full analysis of these facts for now.

It is necessary for Harley's analysis that only P_{HAVE} idioms are compatible with *get*, and so if her analysis is to be extended to account for this data, we must reject the idea that the Scottish Gaelic PDC cannot have P_{HAVE} as a substructure.

4 The indirect objects are PPs

One possible objection to the preceding reasoning is that the IO PPs in the crucial examples illustrating caused possession interpretations are not true PPs, but rather some sort of dative-marked DPs with the *do* being a case marker rather than a semantically non-trivial adposition. This would allow the relevant interpretations to be tied to a specific argument structure frame, and it would be similar in spirit to the analysis of "unexpected" PDCs in English in Bruening (2010b), where the PDC works exceptionally with a heavier or *wh*-extracted IO DP, as documented extensively in work by Joan Bresnan and colleagues (see e.g., Bresnan et al. 2007; Bresnan 2007; Bresnan & Hay 2008; Bresnan & Nikitina 2009).

(36) a. * The stench gave a headache to me.
b. A stench or smell is diffused over the ship that would give a headache to the most athletic constitution.

(37) a. * Nixon's behavior gave an idea for a book to Mailer.
b. Nixon's behavior gave an idea for a book to every journalist living in New York City in the 1970s. (Snyder 2003: 35)

Bruening argues that these PDCs are actually "true" DOCs "disguised" as PDCs, as they are derived from DOC frames but the IO is extraposed rightward and then appended with a dummy *to*. Bruening calls this "R-dative shift".

Putting to one side the question of whether this analysis of these unexpected PDCs in English is feasible, I can demonstrate that it won't work very well at all for Scottish Gaelic, since the language has some reliable diagnostics for PP-hood that diagnose these IOs as true PPs.

4.1 Clefting

One source of evidence for the prepositional status of IOs comes from clefting, which is highly productive in Goidelic. As described in detail by Adger (2011), clefts in Scottish Gaelic involve two distinct "augment" elements, *e* 'it' and *ann* 'in it', which occur between the copula and the cleft pivot. What is important for current purposes is that the augment used depends on the category of the clefted

phrase: if the clefted category is a DP or a (finite or nonfinite) clause, *e* must be used, whereas if the clefted category is a PP (whether adjunct or argument), AdjP, AdvP or an AspP (e.g., a progressive VP), the augment must be *ann*.[10] Some illustrative examples from Adger (2011) are given in (38–42).

(38) 'S {e/*ann} Calum a thug an cat do Mhàiri.
COP.PRES AUG Calum REL give.PST the cat to Mary
"t's Calum who gave the cat to Mary.'

(39) 'S {e/*ann} gu robh e tinn a thuirt mi.
COP.PRES AUG that be.PST e ill REL said I
'What I said was that he was ill.'

(40) 'S {*e/ann} a Ghlaschu a chuir Calum an leabhar.
COP.PRES AUG to Glasgow REL send.PST Calum the book
'It's to Glasgow that Calum sent the book.'

(41) 'S {*e/ann} air a' bhòrd a sheas Lilly.
COP.PRES AUG on the table C.REL stand.PST Lilly
'It's on the table that Lilly stood.'

(42) 'S {*e/ann} breagha a tha i.
COP.PRES AUG beautiful REL be.PRS she
'She's beautiful.'

The examples in (43–48) show that the *ann* augment is also used for the indirect objects in ditransitives of all kinds, including the kind that resisted the PDC frame in English.

(43) 'S {*e/ann} gu Iain a chuir Calum an leabhar.
COP.PRES AUG to Iain REL send.PST Calum the book
'It's to Iain that he sent the book.'

(44) 'S {*e/ann} do Mhàiri a thug Calum an cat.
COP.PRES AUG to Mary REL give.PST Calum the cat
'It's to Mary that Calum gave the cat.'

[10]Very similar facts seem to obtain in Irish, going by the discussion in McCloskey (1984: 469), and the phenomenon also resembles the category-specific changes in the form of the *rannig* element that tracks the fronted category in Breton V2 clauses (Rezac 2004). Like Scottish Gaelic, the Irish and Breton constructions use one form for DPs and another for AdjPs, PPs and at least some VPs. More detailed comparison of the languages is required to establish whether they make the exact same split in all the languages.

(45) 'S {*e/ann} do dh'Iain a thug e ceann goirt.
COP.PRES AUG to Iain.DAT REL give.PST it sore head
'It was Iain that it gave a sore head.'

(46) 'S {*e/ann} do dh'Iain a thug i a' bhròg.
COP.PRES AUG to Iain.DAT REL give.PST she the shoe
'It was Iain that she fired.'

(47) 'S {*e/ann} air Iain a thug i ruith.
COP.PRES AUG on.him Iain REL give.PST she running
'It's Iain that she gave a telling off.'

(48) 'S {*e/ann} às Iain a thug i a' char.
COP.PRES AUG out.of.him Iain REL give.PST she the twist
'It's Iain that she tricked.'

The augment which normally occurs with DP pivots, *e*, is impossible in all of these cases, and there are no other ways to formulate the cleft.

The use of *ann* with clefted *do*-PPs strongly suggests that the fronted indirect object is a PP and not just a DP with an "inserted" case-like adposition. This argument is lent considerable weight by the fact that Adger (2011) provides a principled explanation for the split augment system of Scottish Gaelic clefts in terms of semantic properties of the categories involved, where it's crucial that PPs and DPs have distinct semantic denotations, and thus that prepositions are not semantically vacuous in the language. Adger's explanation of the cleft paradigm is lost if *do* is taken to be a dummy element that is inserted postsyntactically. This, together with the other arguments just outlined, should give sufficient cause to dismiss such a postsyntactic account of the *do*-phrases in idioms and the like.

One other nuance to mention is that dative subjects in Scottish Gaelic actually behave like DPs with respect to this diagnostic, in that clefting of the dative subjects results in the appearance of a bare DP and the *e* augment. As background, Scottish Gaelic is like Irish (documented in detail by McCloskey 1984) in allowing for a variety of raising constructions where the raised subject is marked by either *do* or *le* 'with', as shown in (49).

(49) a. B' fheàrr leinn fo 'r cois iad.
COP.COND better with.1.PL under our feet them
'We'd prefer them under our feet.'

b. B' fheàrr dhomh bhi 'g obair air a' bhuntàta.
COP.COND better to.1SG be ASP work.INF on the potatoes
'I had better work on the potatoes.'

c. Bu chòir do Mhàiri a bhidh ann.
COP.COND duty to Mairi PRT be.INF there
'Mary ought to be there.' (Adger 2022)

In Scottish Gaelic, clefting of the *do*-subjects results in dropping of the P and the use of the *e* augment, as shown in (50):

(50) 'S {e/*ann} Màiri a bu chòir a bhidh ann.
COP.PRES AUG Màiri REL COP.PRES duty PRT be.INF there
'It's Mary that ought to be there.'

As Adger (2022) notes, this is potentially related to the fact that Scottish Gaelic also shows case connectivity in clefts with respect to the genitive. That is, when a genitive DP (for instance the postverbal object of a verbal noun in the progressive) is clefted, it is realized in the unmarked case in its derived position, and genitive is impossible.

(51) a. Bha thu a' geàrradh { na craoibhe / *a' chraobh }
be-PAST you prt cut.VN the-GEN tree-GEN the tree-NOM
'You were cutting the tree.'

b. 'S e { *na craoibhe / a' chraobh } a bha thu a'
COP.PRES AUG the tree.GEN the tree.NOM C.REL be.PAST you PRT
geàrradh
cut.VN
'It's the tree that you were you cutting.' (Adger & Ramchand 2005)

Thus, it's in principle possible for the *do* to be dropped, under circumstances that warrant closer attention (see Thoms 2024). However, this never happens with IOs, which strongly suggests they are true PPs and not dative DPs.

4.2 Pied-piping with inversion

Another strong diagnostic of PP-hood in Scottish Gaelic is "pied-piping with inversion". In this construction, which is found in Irish as well as various other predicate-initial languages (Broadwell 2006), a *wh*-fronted DP that originates as the complement of a preposition occurs sentence-initially and is immediately followed by the selecting P, which itself is 3SM inflected (which is represented as 'P-him' in the same way that it is under P-stranding; see 52 and 53).

(52) Cò ris a bhruidhinn thu?
who with-him C.REL speak.PST you
'Who did you speak to?'

(53) Dè am bòrd fon an do chuir thu an cathair?
What the table under-him C.REL put.PST you the chair
'Which table did you put the chair under?'

Pied piping with inversion occurs with all prepositions as a natural class in Scottish Gaelic, and it can be analyzed (as Adger 2022 does) as involving movement to Spec,PP (understood as the escape hatch of the PP; Abels 2003) followed by pied-piping of the containing constituent.

(54) $[_{CP}\ [_{PP}\ \text{whP}\ [_{P'}\ P\ \langle wh\rangle\ ]]\ [\ C\ \ldots\ ]]$

For similar proposals for Mayan, see Heck (2009: 90) and Cable (2010: 186ff). Such an analysis necessitates a further analysis where the P is itself part of the syntactic representation and not just some post-syntactically inserted formant, because it provides a specifier that can serve as a target for phrasal movement. A post-syntactically inserted case marker could not feed such processes.

To come back to the matter at hand, pied piping with inversion is possible with IOs of ditransitives, including the crucial cases discussed above, as shown in (55–59).

(55) Cò dha a thug i ball?
who to.him C.REL give.PST she ball
'Who did she give a ball to?'

(56) Cò dha a thug e ceann goirt?
who to.it C give.PST it sore head
'Who did it give a sore head?'

(57) Cò dha a thug i a' bhròg?
who to.him C give.PST she the shoe
'Who did she fire?'

(58) Cò air a thug i ruith?
who on.him REL give.PST she running
'Who did she give a telling off?'

(59) Cò às a thug i a' char?
who out.of.him REL give.PST she the twist
'Who did she trick?'

Since pied piping with inversion picks out prepositions as a natural class, it makes sense if the *do* in (56–57) is also a preposition.

Comparing again with dative subjects in Scottish Gaelic, the pattern is the same: while *wh*-movement of dative subjects in Scottish Gaelic involves dropping the preposition altogether (60, but see Thoms 2024 for more discussion), this is wholly impossible with IOs of PDCs of any kind. (61) provides one illustration that represents what is found more generally with PDCs.

(60) a. Is fheàrr *(dha) falbh
COP.PRES better to.him leave.VN
"He had better leave"
b. Cò (*dha) as fheàrr falbh?
who to.him REL.COP better leave.VN
'Who had better leave?' (cf. Adger 2022, 20)

(61) Cò *(dha) a thug i ball?
who to.him C.REL give.PST.IND she ball
"Who did she give a ball to?"

If the droppability of the P with dative subjects is a signal of its case marker status, the impossibility of dropping the P under *wh*-movement is further evidence that the P in IOs is not a case marker.

5 Taking stock

So far, I have shown that Scottish Gaelic PDCs have a broader range of uses than English PDCs, given that they are compatible with caused possession interpretations and various other ditransitive frames where the IO is not a goal PP. This suggests that there is no deep semantic incompatibility between the grammatical ingredients of a PDC and the ingredients that go into making a caused possession ditransitive. If the caused possession interpretation is given by an underlying possessive small clause, as on Harley's account, it must be possible to map this small clause kernel onto a PDC in Scottish Gaelic. This will have to be done without recourse to an analysis where the Ps in the PDCs are just post-syntactically inserted case markers, as we have seen substantial evidence

of their syntactic and semantic activity in Scottish Gaelic PDCs. Given that there is a way to map a possessive small clause onto a PDC frame in Scottish Gaelic, the apparent failure of such a mapping in English – as evidenced by the persistence of Oehrle effects – will require an alternative explanation that appeals to syntactic differences between the languages.

In many ways, much of the original data from English already supported this conclusion. As noted at the beginning of Section 4, caused possession is often possible when the IO is heavy. Also as noted above, Bruening (2010b) proposed an account of these cases that put them to one side. However, Bruening's account was quite ad hoc, requiring preposition insertion in a syntactically very narrowly defined context, and it put to one side the evidence that the Ps in question in English are also quite like genuine prepositions (see Bruening 2010b, 291 fn.5). A more explanatory account of these prosodic weight effects is required, although it's not clear that the pursuit of such an account favours the two-frame approach over the derivational approach. As already discussed, the idea that the semantics filters out certain PDCs has been shown to have little explantory power. In working out an account of the prosodic weight effects in English, one also needs to reckon with the fact that some alternation-resistant ditransitive frames are either consistently resistant to the PDC frame, as with (62), or variably so, as with (63–64). In my experience, (62c) is rejected by everyone regardless of weight manipulation, whereas (63) and (64c) are rejected by many, but not all, speakers (e.g., Bruening 2010b: 295 mentions that "some speakers do not accept R-dative shift at all with *spare*").

(62) a. The recession cost my grandfather a raise.
b. *The recession cost a raise to my grandfather.
c. *The recession cost a raise to everyone in my grandfather's generation. (cf. Rappaport Hovav & Levin 2008, 144)

(63) *Lipson's textbook taught Russian to thousands of students around the world. (Takagi 2009, 70)

(64) a. Let's spare John that ordeal.
b. Let's spare that ordeal to John.
c. *Let's spare that ordeal to all of the undergraduates in our first year class.

Bruening's R-dative shift rule inserts a preposition to mark an extraposed DP in its derived position. How can a rule which marks the IO in a derived context post-syntactically be made sensitive to the lexical content of the main predicate?

Another fact to mention is that several of the sources of evidence from the literature for a possessive core for DOCs get the same results for English PDCs. Beck & Johnson (2004) argued that the ambiguity between repetitive and restitutive readings of *again*, (65), diagnosed a possessive substructure within DOCs. And yet it seems to have been lost in the literature that the authors noted the same ambiguity with English PDCs (see also Ormazabal & Romero 2023, 23–24).[11] In fact, the identical ambiguity holds in Scottish Gaelic PDCs as well, as (67) shows.[12]

(65) Thilo gave Satoshi the map again.
= Thilo gave Satoshi the map, and Thilo has done that before. *repetitive*
= Thilo gave Satoshi the map, and Satoshi had had the map before. *restitutive*

(66) Thilo gave the map to Satoshi again. (both readings, same as 65)

(67) Thug Iain an leabhar ud dhomh a-rithist.
gave Iain the book that to.me again
'Iain gave that book to me again.' (both readings)

If such facts are to be captured with possessive argument structure components, then even English PDCs must have such a core as well.

One final fact about English ditransitives to mention is the fact that not all caused possession ditransitives are as incompatible with the PDC as others. That is, there are exceptions to the generalizations behind the Oehrle effects and associated points of variation. Pesetsky (1995) already noted some ditransitives involving clear caused possession in which there's no clear contrast between the PDC and DOC, as shown in (68) and (69).

(68) a. The conversation with Sue handed Bill a golden opportunity.
b. The conversation with Sue handed a golden opportunity to Bill.

(69) a. Talking to Bill for just a few seconds would have brought Sue a modicum of happiness.
b. Talking to Bill for just a few seconds would have brought a modicum of happiness to Sue.

[11] Bruening (2010a: 550–555) includes an extensive discussion of how to deal with the *again* facts in his applicative version of the analysis of DOCs but does not address the fact that these readings are available with PDCs.

[12] Like with most adverbs in Gaelic, the most natural position for the adverb is at the end of the sentence, although the PP can marginally follow the adverb as well. Postposing of a pronominal DO (or indeed leaving it in situ) does not affect the available interpretations either.

It seems relevant that both of these involve quasi-idiomatic DOs and verbal roots that convey a motion component that's not part of *give*. Even if they are treated as exceptions on this basis, they still demonstrate that *to*-phrases in PDC need not always be goals. This, in turn, suggests that what rules out PDCs in the Oehrle effect cases is unlikely to be some strong requirement that the IOs be goals.

Where does all of this leave us? The Scottish Gaelic facts show that the failures of the PDC in the Oehrle effects cases in English cannot be explained by semantic incompatibility between the PDC frame and the possessive semantics of certain ditransitives. The English facts just outlined indicate that many standard PDCs may well have a possessive semantics. The exceptions to Oehrle effects in English also suggest that *to*-PPs in English ditransitives need not always be goals. What, then, will explain the difference between Scottish Gaelic and English with respect to the Oehrle effects cases? In what follows I pursue the hypothesis that the relevant factor is the interpretation of the external argument, which in the cases in question is a non-agentive causer.

6 Towards a syntactic account

In this section I outline a syntactic account of the distribution of Oehrle effects in terms of the core postulates of a version of the derivational approach, where the crucial difference between the languages resides in the range of available object positions and their interaction with causer subjects, which I argue are base-generated lower than agentive subjects.

6.1 A derivational approach to the dative alternation

I propose an analysis in terms of the derivational approach of the dative alternation from Collins (2024) as applied to Scottish Gaelic (see also Thoms 2024). First, I address the analysis for English – focusing especially on the Oehrle effects. The guiding intuition is that DOCs and PDCs differ in aspects of their object shift syntax, whereby the DO of PDCs typically ends up particularly high. For Collins' analysis of the dative alternation in English, the DOC is analyzed as in Figure 2 (page 164). This represents the (UTAH-compliant) base order from which the PDC is derived, and which involves VP-fronting to the specifier of a *v*P-internal VoiceP projection, as illustrated in Figure 3.

For his analysis of English PDCs, Collins posits that the smuggled DO undergoes object shift to some higher position where it c-commands the IO. This is higher than the licensing position of either object in a DOC. Collins (2024) analyzes the IO as a KP occupying the specifier of an ApplP projection, but I depart

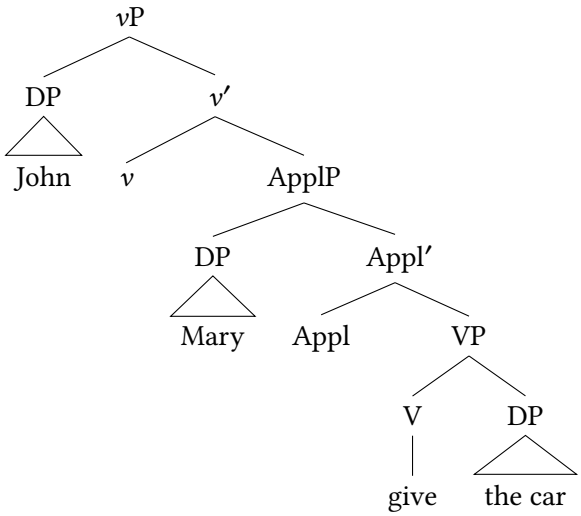


Figure 2: Tree for the DOC

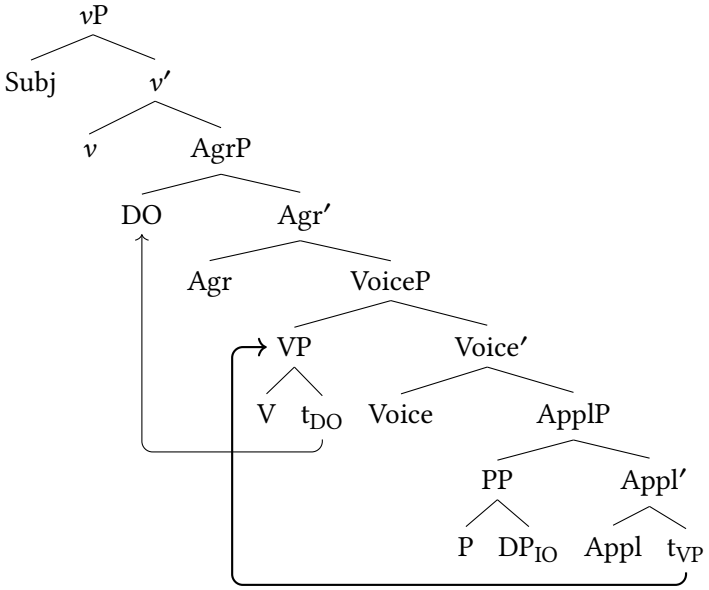


Figure 3: Tree for the PDC

from his analysis because i) we have already seen that the IO is in fact a PP in Scottish Gaelic, and ii) the aim here is to provide a unified account of PDCs in both languages. We can take the ApplP to encode the possessive semantics of both ditransitive constructions, although see Thoms (2024) for more on that matter.

6.2 The position of the external argument and object shift

With this version of Collins' analysis in hand, I add two additional ingredients in order to build an account of Oehrle effects. The first ingredient, which builds on work on *v*/Voice and differences between causer and agent subjects (Pesetsky 1995, Folli & Harley 2005, 2006, 2007, Schäfer 2008, Bjorkman 2018), is the claim that causer subjects and agent subjects are merged in syntactically distinct positions. This is relevant for the Oehrle effects because it seems that a descriptively adequate account of those effects can be stated in terms of the external argument's role, rather than in terms of possession relations holding between the internal arguments. Consider again some of the famous *Mailer* examples, (70) and (71), from Oehrle (1976):

(70) Nixon gave Mailer a book.
  a. = Nixon handed Mailer a book.
  b. = The topic of Nixon (his behavior) gave Mailer an idea for a book

(71) Nixon gave a book to Mailer.
  a. = Nixon handed Mailer a book.
  b. ≠ The topic of Nixon (his behavior) gave Mailer an idea for a book

On the Harley (2002) analysis, the (b) reading is ruled out with (71) because the PDC requires a location argument as IO, and that's not semantically compatible with this *idea-for-a-book* IO. (A similar explanation holds for the related proposals in Bruening 2010a and Harley & Jung 2015.) An alternative description, however, is that it's the external argument's role that matters, given that this is another difference between the (a) and (b) readings: for (a), the subject is (or can be) an agent, but for (b) it is a causer. As far as I can tell, this accounts for the full set of data.

I propose a variant of the analysis in Folli & Harley (2005; 2006; 2007), who claim that there are two different "flavors" of *v* that introduce external arguments with different thematic roles: v_{do}, which introduces agents, and v_{cause}, which introduces causers. Adopting an approach like Ramchand (2008), and departing

somewhat from Folli & Harley (2005; 2007),[13] I propose specifically that the two v heads are merged in different positions, whereby v_{cause} is lower in the clausal spine than v_{do} as in Figure 4.

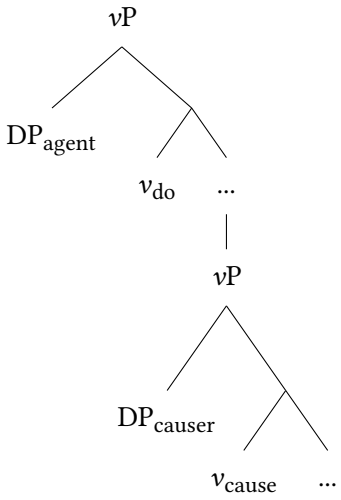


Figure 4: v_{cause} and v_{do} are distinct

With such an analysis in hand, the Oehrle effects in English can be accounted for by claiming that v_{cause} is compatible with DOCs but not with PDCs. In Scottish

[13]Folli and Harley assume that the two flavors of v are merged in the same position. However, they also propose that the two flavors are syntactically distinct, in that v_{cause} may select for a small clause complement and show intransitive verbal morphosyntactic properties, while v_{do} only ever selects verbal structure associated with a transitive clause. This is motivated by the fact that non-agent subjects are possible with verbs of consumption in combination with a resultative small clause, with a concomitant switch to intransitive syntax in the Italian counterparts.

(i) The washing machine chewed *(up) the clothes. (Folli & Harley 2005: 104)

(ii) a. *L'inflazione ha (ri)succhiato i risparmi
the.inflation has suck.PST the savings

b. L'inflazione si é (ri)succhiata i risparmi (Folli & Harley 2005: 104)
the.flation REFL is suck.PST the savings

This is tantalisingly similar to the proposal here, since on my analysis v_{do} selects an object shift-licensing position (which might tie directly to transitive verbal morphosyntax), whereas v_{cause} is lower, potentially selecting syntactic material pertaining to the licensing of small clause substructures. I leave it to future research to explore this connection with Folli and Harley's work more carefully.

Gaelic, the restriction against using v_{cause} with PDCs does not apply. On such an account, PDCs and DOCs are both, in principle, compatible with a possessive core, where the IO is a possessor and the DO is a possessor.

The question, then, is why the lower v_{cause} projection would be incompatible with the PDC frame in English but not in Scottish Gaelic, and, indeed, why v_{cause} would always be compatible with the PDC. This is where the second ingredient comes in. In line with much work on object licensing (e.g., Johnson 1991; Collins & Thráinsson 1993), I propose that DOs must be licensed in AgrP-type projections in the extended verbal projection by means of object shift. This object shift may differ across languages, for instance by targeting different positions in different languages. The crucial claim for my analysis is that in the case of English PDCs, the high object shift position for DOs, which I call simply HighAgrP, is above the merge position of v_{cause}, as shown in Figure 5 (page 168). Recall from Figure 3 above that the launching position for object shift is a VP that has been moved over the IO in Spec,ApplP projection (an instance of smuggling).

This gives us a way to explain the ban on causer subjects in PDCs: the sequence of moving the object to HighAgrP and then moving the causer to Spec,TP is ruled out by Relativized Minimality, which I assume to place severe limitations on A-movement of DPs over one another. Thus, the DO in a PDC derived by smuggling will be required to undergo object shift in order to be licensed. However, it is incapable of doing so in English when a low causer subject is present, because the causer will intervene between the smuggled object and HighAgrP. If the subject argument is an agent, this will be base-generated higher, and so there will be no intervention effect impacting upon object shift of the DO. To make the difference clearer, (72) and (73) schematize the two alongside one another.

(72) $[_{\text{HAgrP}}\ \text{DP}_{\text{DO}} \ldots [_{v\text{P}}\ \text{DP}_{\text{causer}}\ v_{\text{cause}} [_{\text{VP}}\ \text{V}\ \text{t}_{\text{DO}}\ ]\ [_{\text{ApplP}}\ \text{DP}_{\text{IO}} \ldots \text{t}_{\text{VP}} \ldots]]]$

(73) $[_{v\text{P}}\ \text{DP}_{\text{agent}}\ v_{\text{DO}} \ldots [_{\text{HAgrP}}\ \text{DP}_{\text{DO}} \ldots [_{\text{VP}}\ \text{V}\ \text{t}_{\text{DO}}\ ]\ [_{\text{ApplP}}\ \text{DP}_{\text{IO}} \ldots \text{t}_{\text{VP}} \ldots]]]$

No Minimality problem will arise with agent subjects, so they will be compatible with the PDC frame. As for the DOC, I will assume that the DO is licensed lower than in PDCs, within the complement domain of ApplP, although I will not speculate on the specifics here (see Haddican & Holmberg 2019; Holmberg et al. 2019 for some recent ideas about object licensing in DOCs).

To address the situation in Scottish Gaelic, I claim that Scottish Gaelic PDCs have broadly the same derivation (see Figure 6, page 168), and I speculate that

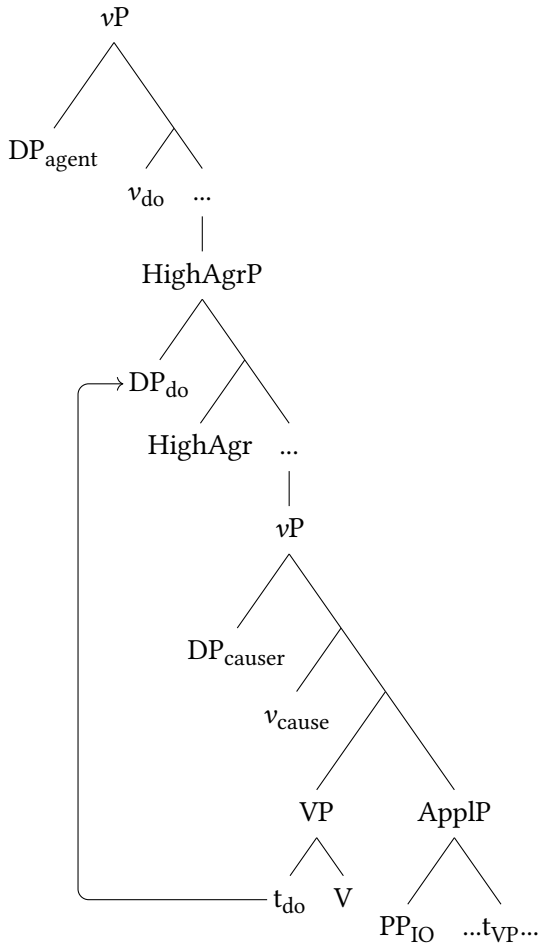


Figure 5: HighAgrP

DOCs are unavailable for the banal reason that DOCs involve marking the IO with a silent preposition (e.g., Emonds & Whitney 2006); silent prepositions are not generally available in these languages because prepositions are always inflected and zero heads cannot bear inflection. As for why the Minimality issue does not arise in the derivation of PDCs with causer subjects, I claim that this is because the DO has a licensing position within the verbal constituent that fronts to the left of the IO in this language, and so object shift to HighAgrP is not required for object licensing.

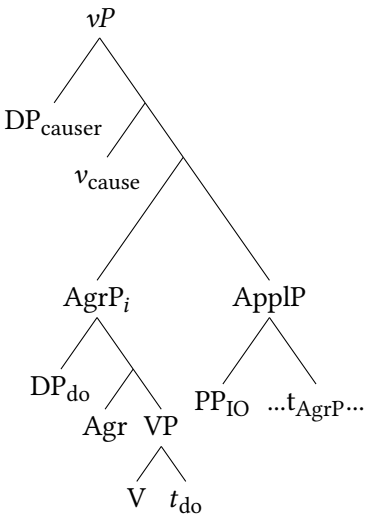


Figure 6: Scottish Gaelic PDCs

If Scottish Gaelic DOs can be licensed low in this projection, with no need to shift into the higher AgrP projection, then no intervention is expected with causer subjects. The claim is that English lacks such an option.

The two postulates here – that causers and agents are merged at different heights, and that the languages differ with respect to the availability of an additional low object licensing position in PDCs – are in need of independent support. I provide some such evidence below.

6.3 Object position differences

There are a few reasons to believe that Scottish Gaelic has a range of lower object positions available for DOs in its ditransitives. Before getting to those arguments, first I should establish some general points about Scottish Gaelic and English object positions more generally.

To begin with, there is ample evidence that objects undergo an obligatory object shift in Scottish Gaelic. Consider word order in nonfinite clauses, in which DOs always occur to the left of the nonfinite verb.

(74) Bu toil leam na leabhraichean a leughadh.
COP.COND desire with.me the books PRT read.INF
'I would like to read the books.'

I assume that the DO always raises to such a position, in finite and nonfinite clauses, and that the only difference between finite and nonfinite clauses is that the verb raises to some very high position in finite clauses.

That the DO has raised to its surface position, and is not just base-generated there in an underlying OV structure, is indicated by the fact that the DO can strand a floating quantifier to the right of the verb, as in (75). In PDCs, the quantifier stranded by a DO comes between the verb and the IO PP, which always occur to the right of the nonfinite verb, as in (76).[14]

(75) Bu toil leam **na leabhraichean** a leughadh **uile**.
COP.COND desire with.me the books PRT read.INF all
'I would like to read all the books.'

(76) Bu toil leam **na leabhraichean** a thoirt **uile** do dh'Iain.
COP.COND desire with.me the books PRT give.INF all to Iain
'I would like to give all the books to Iain.'

This point is strengthened further by the fact that certain objects, in particular measure phrases that occur as objects of predicates like *cosg*, do not shift. Adger (1994) provides an explanation for why these objects in particular would not undergo object shift.

(77) Feumaidh a' cho-labhairt cosg trì mìle not.
must the conference cost.INF three thousand pound
'The conference has to cost three thousand pounds' (Adger 1994: 110)

Beyond cases such as (77), raising of DOs to the left of the verb in infinitives is absolutely obligatory, whatever the thematic role of the subject may be. This all fits with the analysis in Figure 6, where there is object shift within the VP that fronts. However, it is also compatible with an analysis where the object shifts to HighAgrP, because the two positions cannot be distinguished by their surface position.

I would like to argue that if the analysis where Scottish Gaelic objects are licensed within the VP is adopted (but English objects are licensed higher) then there is a principled explanation for some interesting differences between English and Celtic PDCs with respect to c-command asymmetries. In Celtic, "backwards" binding from the IO into the DO is available very freely, seemingly for all speakers, as the examples in (78–80) indicate. Note that this includes "high" applicatives such as (80) as well as classic "low" applicatives such as (78).

[14]For more discussion of quantifier float in Irish and Scottish Gaelic, see Maki & Ó Baoill (2008; 2011); McCloskey (2017); Thoms (2024).

(78) Thug mi a$_i$ pheann-fhèin do a h-uile gille$_i$.
give.PST I his pen-own to every boy
'I gave every boy his own pen.'

(79) Sheall Màiri a$_i$ leabhar(-fhèin) dhan a h-uile ùghdar$_i$.
show.PST Mary his book(-own) to every author
'Mary showed every author his book.' (Jung et al. 2012, 11)

(80) Chàraich Màiri a$_i$ choimpiutar-fhèin airson gach / a h-uile caraid$_i$
fix.PST Mary his computer-self for each every friend
'Mary fixed his$_i$ computer for every friend.' (Jung et al. 2012, 14)

"Forward" c-command from the DO into the IO, on the other hand, is only available to a subset of speakers. Jung et al. (2012) report forward binding is unavailable in (81) for the speakers they consulted, whereas it was possible for my consultant.

(81) %Sheall Màiri a h-uile leabhar$_i$ dha (h-)$_i$ ùghdar.
show.PST Mary every book to its author
'*Mary showed its$_i$ author [every book]$_i$.' (Jung et al. 2012, 11)

The situation where backwards c-command is available but forward c-command is not is also seen in ditransitives in Austronesian languages such as Malagasy (Pearson 2000) and Mandar (Dan Brodkin, p.c.). The variability in Scottish Gaelic must derive from some additional ingredient that is acquired by some but not all speakers.

More or less the opposite picture is found in English. With English PDCs, all speakers allow forward binding from DOs into IOs, as has been well-documented since Barss & Lasnik (1986; see also Larson 1988). See the examples in (82).

(82) a. John gave [every dog]$_i$ to its$_i$ owner.
b. Jane sent [every boy]$_i$ to his$_i$ parents.
c. Kim fixed [every bike]$_i$ for its$_i$ owner.

On the other hand, a subset of speakers allow "backwards" c-command from IO into DO. Collins (2024) reports that it is possible for him and many speakers, and gives citations to various places in the literature where it's been reported that backwards binding is at least marginally possible in examples such as (83).[15]

[15]For some speakers, backwards binding is improved a lot with a heavier IO (Craig Sailor, p.c.):

(83) John gave/sent his$_i$ paycheck to [every worker]$_i$.

But some report these readings to be unavailable; see for instance Larson (1988: 339), who reports such examples to be "??" and points to various other diagnostics that seem to suggest the DO asymmetrically and consistently c-commands the IO in PDCs, as with weak crossover in questions such as in (84).

(84) *[Which worker]$_i$ did you send his$_i$ paycheck to?

It seems this is a point of variation, and it is the opposite of that seen in Scottish Gaelic.

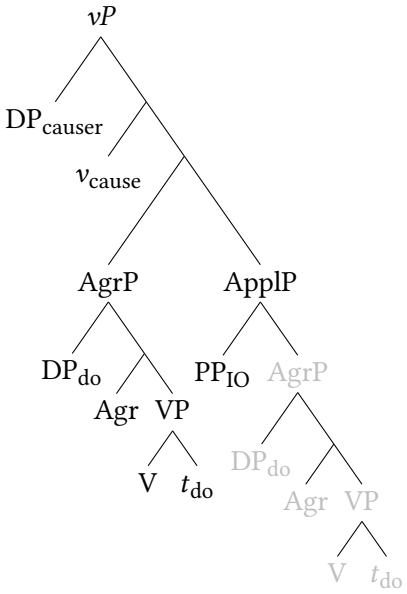


Figure 7: VP fronting in PDCs

This picture can be understood in terms of the VP-fronting analysis of PDCs combined with the object positions posited in Figure 7. Scottish Gaelic is ac-

(i) John gave his$_i$ paycheck to [every man who came to the office to complain]$_i$.

One possible analysis of the variation with backwards binding in English is that the IO only c-commands into the DO if the IO is extraposed to a position above HighAgrP, with no role to play for reconstruction of the DO at all. If this were right, the variation could be attributed to the fact that some speakers accommodate such an analysis for less heavy objects (such as 83) more readily than others. Clearly, it would be desirable to link this to an analysis of heaviness effects elsewhere, for instance those discussed by Bruening (2010a).

counted for if I add that objects can be licensed within the fronted verbal constituent, with reconstruction of that verbal constituent delivering "backward" c-command from the IO into the DO by default (as Pearson 2000 does for Malagasy).

As for the speakers who allow forward c-command, there is the possibility of shifting the object to HighAgrP, string-vacuously, after which the DO would asymmetrically c-command the IO. The point of variation here is whether the speakers have this string-vacuous object shift rule; this is arguably a plausible source of variation, because the positive evidence for such a rule for the learner would typically be quite indirect.

The inverse situation would hold in English: for speakers who allow backwards binding, there must be the possibility of licensing the object within the fronted VP; this feeds reconstruction and, thus, backwards binding. For those speakers who don't allow backwards binding, this low licensing position would be unavailable; thus, object shift to the higher position would be mandated in all PDCs. That there is object shift of the DO in PDCs in English is supported by quantifier float facts, where the DO in a PDC can strand a quantifier to its right and before the IO. See the examples in (85).

(85) a. I gave the cards all to their recipients.
b. He read the books both to someone famous. (Janke & Neeleman 2012: 164)

If English had the same range of object positions as Scottish Gaelic, it would be difficult to understand the differences observed here. By analyzing the availability of backwards c-command as indicative of a low licensing position for the DO, the differences are explained succinctly and with an analysis that dovetails with the account of Oehrle effects outlined earlier.

Finally, the analysis just outlined may also provide a way to understand the exceptional cases from Pesetsky (1995), repeated here as (86) and (87):

(86) a. The conversation with Sue handed Bill a golden opportunity.
b. The conversation with Sue handed a golden opportunity to Bill.

(87) a. Talking to Bill for just a few seconds would have brought Sue a modicum of happiness.
b. Talking to Bill for just a few seconds would have brought a modicum of happiness to Sue.

Exceptional cases may involve the low AgrP selecting for specific verbs, something that would need to be learned on the basis of input on a case-by-case basis. Pesetsky notes (p.319, n. 171) that not all speakers accept the PDCs in the (b) examples here, so there's more empirical work to be done to figure out what's going on. It seems more likely that the variation is syntactic, interacting with the lexical semantics of the predicate, rather than something that's purely a matter of the incompatibility of the ingredients of PDCs and caused possession.

6.4 Causer subjects are lower

The second ingredient of the analysis, that causer subjects are lower, can be motivated empirically in a few ways.

First, the idea that agents are base-generated higher than other external arguments is insightfully exploited by Bjorkman (2018) to provide an account of the interaction between agentivity and ergative marking on unergative subjects in "split-S" ergative systems such as that of Hindi-Urdu, where certain non-agentive subjects show up without ergative marking. Bjorkman (2018: 343) provides an analysis of such a system that makes crucial use of the idea that nonagentive subjects may originate in a lower position than agentive subjects as part of a wider analysis of the various factors that impact upon split ergativity.

Second, Pesetsky (1995) notes a range of facts that point to a low position for causer subjects. In particular, he points out that in ditransitives with causer subjects, the kind that are implicated in Oehrle effects, the IO can "backwards" bind into the subject, both for anaphor binding purposes, as shown by (88–89) from Pesetsky (1995: 201), and also variable binding purposes, (90). Such backwards binding is not possible with agentive subjects, as shown in (91–92).

(88) Each other's remarks gave John and Mary a book. (Pesetsky 1995: 201)

(89) Those books about himself taught Bill the meaning of caution. (Pesetsky 1995: 201)

(90) His/her$_i$ first qualifying paper gave [every student]$_i$ a headache.

(91) *Each other's parents gave John and Mary terrible gifts.

(92) *His$_i$ parents gave [every boy]$_i$ terrible gifts.

Pesetsky captures the backwards binding with causer subjects in terms of a very low position for causer subjects, along with a rather radical theory of phrase structure, which I will not get into here. An alternative analysis is that causer

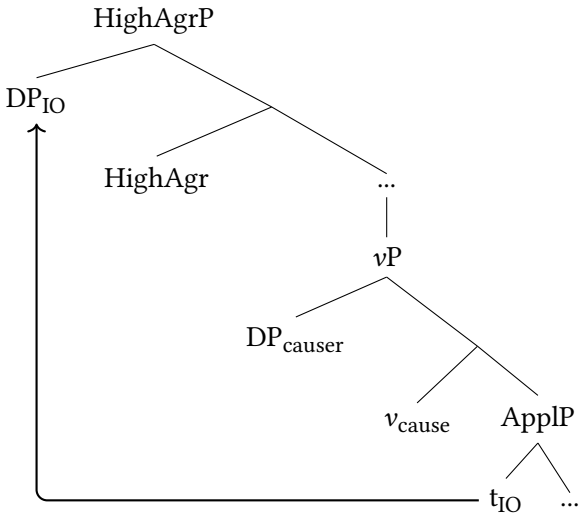


Figure 8: IO shift

subjects are below HighAgrP as proposed above, and the IO can be object shifted to Spec,HighAgrP, from where it can bind into the causer subject, as in Figure 8.

Reconstruction of the raised subject to this low Spec,v_{cause} position should thus allow for "backwards" binding from the object into the subject, whereas it will continue to be impossible with agents, which are too high for any shifted objects to bind into.

The analysis just outlined may seem to be in direct conflict with the analysis I outlined in Figure 5 above for the core Oehrle effects cases: why does the causer subject block object shift in Figure 5 but not in Figure 8? My proposal is that the relevant difference is that in Figure 8 the two DPs differ in terms of animacy: in configurations such as (88–90), the IO is animate, whereas the causer is inanimate (or can be analyzed thus, unlike with an agent[16]), and this difference allows for A-probes to discern between the two and, thus, for A-locality issues to be circumvented with structures such as Figure 8. On the other hand, in Figure 5

[16]Examples such as the following would seem to pose a problem for this analysis:

(i) Her$_i$ political enemies gave [every journalist]$_i$ the impetus to keep investigating.

I assume that the subject here, when analyzed as a causer, can be coerced as an inanimate with a meaning akin to "the idea/thought of his enemies", possibly with a silent noun embedding the overt nominal. I presume a similar analysis would obtain for cases such as *we were discussing John*.

above, a structure representing the core Oehrle effect cases, both the external argument and the shifted DO are inanimate (**the noise gave a headache to me*), and so Relativized Minimality would not be able to discern between the two in a way that resolves the locality issue.

The idea that subjects and objects would interact with respect to their relative animacy is not a new one of course; some languages restrict certain combinations of subjects and objects according to animacy hierarchies. (93) is an example from Chuj (from Deal & Royer 2023):

(93) a. Ix-y-il nok' chan winh winak.
PFV-A3-see CLF snake CLF man
'The man saw the snake.'

b. *Ix-y-il winh winak nok' chan.
PFV-A3-see CLF man CLF snake
Intended: 'The snake saw the man.'

Therefore it is plausible, if somewhat novel, in the landscape of English to claim that animacy features would be involved in a Minimality calculus impacting upon object shift. This analysis brings animacy back into the fold as a means to understanding effects on the dative alternation, but by a rather different route than what was taken by Oehrle (1976) and Harley (2002), where animacy determined whether a given element could be a possessor.

On the account here, animacy doesn't play a direct role in determining if a given argument can receive one theta role or another, but rather it determines the range of possible crossing paths, and in turn, the range of possible ditransitive derivations allowing backwards binding. This is important for a few reasons. Harley & Jung (2015) make it clear that animacy is not required for a given argument to appear as a possessor IO in a DOC; rather, all that matters is that the related possessive predication is well-formed. The examples in (94–95) from Harley & Jung (2015) demonstrate this.

(94) a. #The advertiser gave the car a flyer.
b. #The car has a flyer.

(95) a. The painter gave the house a new coat of paint.
b. The house has a new coat of paint.

Similar remarks hold for external arguments. Folli & Harley (2005; 2006; 2007) show that agents can be inanimate so long as they have "teleological capability"; for example, computer programs, machines or other inanimates can receive

agent roles, such as in *the kettle whistled*. Thus, it is expected that there is no strong ban on inanimates as the (agentive) subject of PDCs in English, and that seems correct, as (96) shows.

(96) a. iCal sent the participants a reminder.
b. iCal sent a reminder to the participants.

Thus, animacy may play some role in determining the semantic compatibility of a given DP with a thematic role, but it is not fully determinative of the range of possibilities. On the other hand, animacy on my account does have a role to play in determining the range of possible and impossible derivations where an object crosses the external argument. This possibility arises with causer external arguments, but not with agents, which are merged above HighAgrP and, thus, cannot be crossed by object shift independent of animacy considerations.

7 Conclusion

In this chapter I have argued that Oehrle effects should be explained not as a side effect of distinct semantic interpretations for DOCs and PDCs. The semantic explanation, which had been taken to support the two-base theory of the dative alternation, struggles to account for the fact that there is no such incompatibility with PDCs in Scottish Gaelic. This sent me in search of a syntactic alternative, one which is capable of handling cross-linguistic variation in the attestation of Oehrle effects. The syntactic alternative that I propose appeals to Relativized Minimality and a more articulated syntax for object licensing. Other syntactic analyses may, of course, present themselves, as data is likewise brought into the picture from other languages. The take-home point here, I suggest, is that apparently semantic effects can quite often be given a syntactic reinterpretation, particularly when there are so many syntactic moving parts to take into account. Comparative data is essential to assessing these matters.

Acknowledgments

Many thanks to Gillebrìde MacMillan above all for his time and endless patience working with me on the data that feeds into my work on Gaelic. Thanks also to Andrew Carnie for inviting me to contribute to the series, and to the two reviewers, the other editors, the audience members at the FACL talk (especially Heidi Harley), and also audiences at LSA 2024 and Edinburgh, David Adger, Chris

Collins, Stephanie Harves, David Pesetsky, Richie Kayne, Alec Marantz, Craig Sailor, and David Willis for discussion of issues that fed into this work.

References

Abels, Klaus. 2003. *Successive cyclicity, anti-locality and adposition stranding*. University of Connecticut at Storrs. (Doctoral dissertation).

Adger, David. 1994. *Functional heads and interpretation*. University of Edinburgh. (Doctoral dissertation).

Adger, David. 2011. Clefted situations: A note on expletives in Scottish Gaelic clefts. In Andrew Carnie (ed.), *Formal approaches to Celtic linguistics*, 1–16. Cambridge: Cambridge Scholars Publishing.

Adger, David. 2022. Mind your Ps and Ks! Quirky subjects and nominal licensing. Ms., Queen Mary University of London (and lingbuzz).

Adger, David & Gillian Ramchand. 2005. Merge and move: *Wh*-dependencies revisited. *Linguistic Inquiry* 36(2). 161–193.

Barss, Andrew & Howard Lasnik. 1986. A note on anaphora and double objects. *Linguistic Inquiry* 17. 347–354.

Beck, Sigrid & Kyle Johnson. 2004. Double objects again. *Linguistic Inquiry* 35(1). 97–123. DOI: 10.1162/002438904322793356.

Bjorkman, Bronwyn. 2018. Ergative as perfect oblique. *Syntax* 21(4). 321–361. DOI: 10.1111/synt.12162.

Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Sam Featherston & Wolfgang Sternefeld (eds.), *Roots: Linguistics in search of its evidential base*, 75–96. Berlin: Mouton de Gruyter.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & Harald Baayen. 2007. Predicting the dative alternation. In Gerlof Boume, Irene Krämer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Bresnan, Joan & Jennifer Hay. 2008. Gradient Grammar: An effect of animacy on the syntax of *give* in New Zealand and American English. *Lingua* 118(2). 245–259. DOI: 10.1016/j.lingua.2007.02.007.

Bresnan, Joan & Tatiana Nikitina. 2009. The gradience of the dative alternation. In Linda Uyechi & Lian Hee Wee (eds.), *Reality exploration and discovery: Pattern interaction in language and life*, 161–184. Stanford, CA: CSLI Publications.

Broadwell, George Aaron. 2006. Alignment, precedence, and the typology of pied-piping with inversion. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG06 conference*. CSLI Publications.

Bruening, Benjamin. 2001. QR obeys Superiority: ACD and frozen scope. *Linguistic Inquiry* 32(2). 233–273. DOI: 10.1162/00243890152001762.

Bruening, Benjamin. 2010a. Ditransitive asymmetries and a theory of idiom formation. *Linguistic Inquiry* 41. 519–562. DOI: 10.1162/ling_a_00012.

Bruening, Benjamin. 2010b. Double object constructions disguised as prepositional datives. *Linguistic Inquiry* 41. 287–305. DOI: 10.1162/ling.2010.41.2.287.

Bruening, Benjamin. 2018. Double object constructions and prepositional dative constructions are distinct: A reply to Ormazabal and Romero 2012. *Linguistic Inquiry* 49(1). 123–150. DOI: 10.1162/ling_a_00268.

Cable, Seth. 2010. *The grammar of Q: Q-particles, wh-movement and pied-piping*. Oxford: Oxford University Press.

Collins, Chris. 2024. *Principles of argument structure: A merge-based approach*. Cambridge: MIT Press.

Collins, Chris & Hoskuldur Thráinsson. 1993. Object shift in double object constructions and the theory of case. In Colin Phillips (ed.), *Papers on case and agreement ii*, 131–174. Cambridge: Department of Linguistics & Philosophy.

Deal, Amy Rose & Justin Royer. 2023. Mayan animacy hiearchy effects and the dynamics of Agree. Ms., UC Berkeley.

Emonds, Joseph E. & Rosemarie Whitney. 2006. Double object constructions. In Martin Everaert & Henk van Riemsdijk (eds.), *The Blackwell companion to syntax*, vol. 1, 73–144. New York, NY: Blackwell.

Folli, Raffaella & Heidi Harley. 2005. Consuming results in Italian and English: Flavors of *v*. In Paula Kempchinsky & Roumyana Slabakova (eds.), *Aspectual inquiries*, 95–120. Dordrecht: Springer.

Folli, Raffaella & Heidi Harley. 2006. On the licensing of causatives of directed motion: Waltzing Matlida all over. *Studia Linguistica* 60(2). 121–155. DOI: 10.1111/j.1467-9582.2006.00135.x.

Folli, Raffaella & Heidi Harley. 2007. Causation, obligation, and argument structure: On the nature of little v. *Linguistic Inquiry* 38(2). 197–218. DOI: 10.1162/ling.2007.38.2.197.

Green, Georgia M. 1974. *Semantics and syntactic regularity*. Bloomington, Indiana: Indiana University Press.

Haddican, Bill & Anders Holmberg. 2019. Object symmetry effects in Germanic: Evidence for the role of Case. *Natural Language and Linguistic Theory* 37. 91–122. DOI: 10.1007/s11049-018-9404-5.

Harley, Heidi. 2002. Possession and the double object construction. *Yearbook of Linguistic Variation* 2. 29–68. DOI: 10.1075/livy.2.04har.

Harley, Heidi & Hyun Kyoung Jung. 2015. In support of the p_{have} analysis of the double object construction. *Linguistic Inquiry* 46. 703–730. DOI: 10.1162/ling_a_00198.

Harley, Heidi & Shigeru Miyagawa. 2017. Ditransitives. In Mark Aronoff (ed.), *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press. DOI: 10.1093/acrefore/9780199384655.013.186.

Heck, Fabian. 2009. On certain properties of Pied-Piping. *Linguistic Inquiry* 40(1). 75–111. DOI: 10.1162/ling.2009.40.1.75.

Holmberg, Anders, Michelle Sheehan & Jenneke van der Wal. 2019. Movement from double object constructions is not fully symmetrical. *Linguistic Inquiry* 50(4). 677–722. DOI: 10.1162/ling_a_00322.

Janke, Vikki & Ad Neeleman. 2012. Ascending and descending VPs in English. *Linguistic Inquiry* 43(2). 151–190. DOI: 10.1162/ling_a_00082.

Johnson, Kyle. 1991. Object positions. *Natural Language and Linguistic Theory* 9(4). 577–636. DOI: 10.1007/BF00134751.

Jung, Hyun Kyoung, Andrew Carnie & Heidi Harley. 2012. On the lack of double object constructions in Scottish Gaelic (and Modern Irish). Handout for a talk at New Perspectives on Celtic Syntax, Berkeley.

Kayne, Richard S. 1993. Toward a modular theory of auxiliary selection. *Studia Linguistica* 47(1). 3–31. DOI: 10.1093/oso/9780195102352.003.0007.

Larson, Richard K. 1988. On the double object construction. *Linguistic Inquiry* 19(3). 335–392.

Larson, Richard K. 2017. On "dative idioms" in English. *Linguistic Inquiry* 48(3). 389–426. DOI: 10.1162/ling_a_00248.

Maki, Hideki & Dónall P. Ó Baoill. 2008. The theme goal construction in Modern Irish. *English Linguistics* 25(2). 439–451. DOI: 10.9793/elsj1984.25.439.

Maki, Hideki & Dónall P. Ó Baoill. 2011. The distribution of floating quantifiers in Ulster Irish and its theoretical implications. In *Essays on Irish syntax*, 35–44. Tokyo: Kaitakusha.

Marantz, Alec. 1993. Implications of asymmetries in double object constructions. In Sam A. Mchombo (ed.), *Theoretical aspects of Bantu grammar 1*, 113–151. Stanford University: CSLI Publications.

McCloskey, James. 1984. Raising, subcategorization and selection in Modern Irish. *Natural Language and Linguistic Theory* 1(4). 441–485.

McCloskey, James. 2017. Object positions. In Nicholas LaCara, Keir Moulton & Anne-Michelle Tessier (eds.), *A schrift to fest Kyle Johnson*, 255–270. Amherst, MA: Linguistics Open Access Publications.

McCloskey, James & Kenneth Hale. 1984. On the syntax of person-number inflection in Modern Irish. *Natural Language and Linguistic Theory* 1(4). 487–533. DOI: 10.1007/bf00417057.

Oehrle, Richard. 1976. *The grammatical status of the English dative alternation*. Massachusetts Institute of Technology. (Doctoral dissertation).

Ormazabal, Javier & Juan Romero. 2012. Pps without disguises: Reply to Bruening. *Linguistic Inquiry* 43(3). 455–474. DOI: 10.1162/ling_a_00097.

Ormazabal, Javier & Juan Romero. 2023. In the beginning was a *to*-phrase. Ms., HiTT.

Pearson, Matthew. 2000. Two types of VO languages. In Peter Svenonius (ed.), *The derivation of VO and OV*, 327–363. Amsterdam: John Benjamins.

Pesetsky, David. 1995. *Zero syntax*. Cambridge: MIT Press.

Pylkkänen, Liina. 2002. *Introducing arguments*. Massachusetts Institute of Technology. (Doctoral dissertation).

Pylkkänen, Liina. 2008. *Introducing arguments*. Cambridge: MIT Press.

Ramchand, Gillian. 2008. *Verb meaning and the lexicon: A first phase syntax*. Cambridge: Cambridge University Press.

Rappaport Hovav, Malka & Beth Levin. 2008. The English dative alternation: The case for verb sensitivity. *Journal of Linguistics* 44(1). 129–167. DOI: 10.1017/s0022226707004975.

Rezac, Milan. 2004. The EPP in Breton: An unvalued categorial feature. In Anne Breitbarth & Henk van Riemsdijk (eds.), *Triggers*. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110197433.451.

Richards, Norvin. 2001. An idiomatic argument for lexical decomposition. *Linguistic Inquiry* 32(1). 183–192. DOI: 10.1162/002438901554649.

Schäfer, Florian. 2008. *The syntax of (anti)causatives*. Amsterdam: John Benjamins.

Snyder, Kieran. 2003. *The relation between form and function in ditransitive constructions*. Philadelphia, PA: University of Pennsylvania. (Doctoral dissertation).

Snyder, William & Karin Stromswold. 1997. The structure and acquisition of English dative constructions. *Linguistic Inquiry* 28(2). 281–317. DOI: 10.58837/chula.the.2021.145.

Takagi, Hiroyuki. 2009. On the causal interpretation of the English ditranstive construction. *Bulletin of the School of Literature, Arts and Cultural Studies, Kinki University* 20(2). 59–77.

Thoms, Gary. 2024. Towards a syntactic analysis of preposition insertion. Ms., NYU.

Chapter 5

Some complexities and idiosyncrasies of Welsh consonantal mutation

Maggie Tallerman
Newcastle University

This chapter examines some of the peculiarities and idiosyncrasies of Welsh initial consonantal mutation and is especially targeted at those unfamiliar with the phenomena of mutation in general. I first provide an outline of the main contexts in which mutations occur in Welsh, which include lexical mutation triggers, morphosyntactic triggers and syntactic triggers. I go on to examine some of the complexities of initial mutation, including the fact that not all mutations have a trigger at all, and not all triggers are overtly realized. Mutations of loanwords, both established loans and nonce loans, are also discussed and illustrated.

1 Introduction

This chapter examines some of the complexities and idiosyncrasies found in Welsh initial consonantal mutation. A central goal is to showcase the specific features of Welsh mutation in the spoken language, including more recent developments in colloquial Welsh and in the mutation of loanwords, both established loans and nonce loans. We see that not all mutation in Welsh is triggered by a preceding lexical item: some mutations are syntactically or morphosyntactically triggered; some are not triggered at all; some mutation triggers are not overtly realized; and some lexical items fail to undergo mutation. We examine variation in mutation, including idiolectal and dialectal variation. More formal varieties of Welsh will be referred to when relevant.

The term "consonantal mutation" refers to sets of morphophonological alternations in the initial segments of words or morphemes. In modern Welsh, these

Maggie Tallerman. 2025. Some complexities and idiosyncrasies of Welsh consonantal mutation. In Andrew Carnie, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond & Luis Irizarry (eds.), *Foundational approaches to Celtic linguistics*, 183–208. Berlin: Language Science Press. DOI: 10.5281/zenodo.15654865

alternations are conditioned by factors other than a purely phonological context. Mutations are realized phonologically, but not triggered phonologically. I use the term "trigger" to refer to the lexical item, phrase or (morpho)syntactic context that conditions the mutation, and "target" to refer to the element that undergoes the mutation. In Welsh, mutation is strictly local and only immediately preceding items are possible triggers. As we will see, not all mutation is triggered, and various factors determine whether a predicted mutation occurs in each context. Section 2 briefly outlines the basic facts of Welsh and illustrates some of the main triggering contexts for mutation; it is not intended to be a comprehensive survey of the mutation system. Section 3 examines some further complexities of Welsh mutation, showing that not all mutation triggers are overt, and that not all mutations have a trigger. Section 4 turns to some idiosyncrasies of mutation, showing speaker variation and the effects of mutation in loanwords. Section 5 examines a mutation paradox, and proposes a solution based on the notion of a non-overt mutation trigger.

2 Overview of Welsh mutations, their triggering contexts and targets

2.1 The basic facts of mutation

Welsh has three series of mutations, traditionally referred to as "soft", "nasal" and "aspirate", with the term "radical" referring to the basic lexical form of the word, which is the citation form. In this paper, I focus on soft mutation (SM), by far the most frequent mutation – both in terms of triggering contexts and occurrence in all kinds of data, but also by far the most interesting and diverse of the different types.

As Table 1 shows, only a word with an initial voiceless stop can undergo each of the three possible mutations; nasal mutation applies to all the six stops, but aspirate mutation only applies to the voiceless stops. I follow Hammond (2019) in assuming that phonologically the voiceless nasals are sequences of a nasal followed by [h]: [mh, nh, ŋh]. My examples here are generally given in Welsh orthography, which, as is clear from Table 1, represents phonetic values somewhat differently than in English. (1) illustrates, with the superscript indicating the item that is the trigger for the specific mutation, and denoting which of the three mutation series (soft, nasal or aspirate) is triggered. Here, the triggers are agreement proclitics that occur in appropriate agreement contexts before nouns and non-finite verbs; these small functional elements are very typical of the types of trigger for Welsh mutations:

Table 1: Welsh consonantal mutations: Orthographic [phonetic]

Radical		Soft		Nasal		Aspirate	
p	[p]	b	[b]	mh	[mh]	ph	[f]
t	[t]	d	[d]	nh	[nh]	th	[θ]
c	[k]	g	[g]	ngh	[ŋh]	ch	[x]
b	[b]	f	[v]	m	[m]		
d	[d]	dd	[ð]	n	[n]		
g	[g]	–	zero	ng	[ŋ]		
m	[m]	f	[v]				
ll	[ɬ]	l	[l]				
rh	[r^{h}]	r	[r]				

(1) Radical: *tad* ‘father’
Nasal mutation: *fy/yn^{N} nhad* (1s father) ‘my father’
Soft mutation: *dy^{S} dad* (2s father) ‘your father’
Soft mutation: *ei^{S} dad* (3ms father) ‘his father’
Aspirate mutation: *ei^{A} thad* (3fs father) ‘her father’
Radical form: *ein tad* (1pl father) ‘our father’

As these examples show, all the singular proclitics trigger one of the three different mutation series. None of the plural proclitics trigger any mutation, so are followed by the radical form of the following item.

2.2 Simple mutation triggers

Most contexts for triggered mutation in Welsh fall into one of two categories: first, mutations which are simply triggered by a preceding item with no restrictions, and second, mutations triggered in specific morphosyntactic or syntactic contexts, in which either trigger or target may exhibit various restrictions. In this section, I illustrate the first category, where the mutation is triggered by an immediately preceding, c-commanding lexical item. Generally, these triggers are functional or semi-functional elements such as those in (1), including various agreement proclitics, prepositions, numerals, conjunctions, determiners, complementizers and particles. Example (2) illustrates some typical examples (triggers in bold; radical form shown in parentheses), all showing soft mutation triggered by a specific preceding element. The relevant c-command domain is shown in square brackets:

(2) a. **gan** [flodyn] (*blodyn*)
with flower
'with a flower'

b. **gan** [bedwar blodyn tlws] (*pedwar*)
with four flower pretty
'with four pretty flowers'

c. **dau** [fis] (*mis*)
two month
'two months'

d. **pa** [ddiwrnod] (*diwrnod*)
which day
'which day'

e. te **neu** [goffi] (*coffi*)
tea or coffee
'tea or coffee'

f. **Mi** [brynais i docyn] (*prynais*)
PRT buy.PAST.1S I ticket
'I bought a ticket.'

g. **yn** [ofalus] (*gofalus*)
PRED careful
'carefully'

The contrast between (2a) and (2b) shows that mutation always targets the initial consonant of the first item in the following phrase. So, in both cases the mutation is realized on the initial consonant of the NP: in (2a) the NP consists only of its head N, while in (2b) the NP is *pedwar blodyn tlws* 'four pretty flowers'.

Mutation does not target the head of the following phrase, which may or may not be the initial element. This illustrates the strict locality of all mutation in Welsh. So, if the initial element of a target phrase is a vowel or does not have a mutable initial consonant – which happens often, since there are only nine mutable consonants and the total consonant inventory is larger – then the phrase gives no indication at all of the triggered mutation. For instance (cf., 2b), in *gan saith blodyn* 'with seven flowers', we have a soft mutation trigger, the preposition *gan*, which has as its target the phrase *saith blodyn*. But the numeral *saith* 'seven' does not have a mutable initial, so there is nothing to indicate here that *gan* is a trigger for soft mutation. The mutation crucially does not skip over the numeral to target the head noun: **gan saith flodyn*. If, on the other hand, the phrase is

gan dri chi, 'with three dogs', then *gan* does c-command a following phrase with a mutable initial, [*tri chi*], and triggers soft mutation as predicted: *tri* > *dri*.

2.3 Morphosyntax and syntax

In the second major category for triggered mutation, either target or trigger display some specific morphosyntactic or syntactic restrictions.

2.3.1 Morphosyntactic restrictions

First, we examine morphosyntactic restrictions on the target. The definite article in Welsh triggers soft mutation, but only feminine singular nouns are a target for this mutation (3a), and never plural nouns (3b), or masculine nouns (3c, 3d). (Welsh has two genders, traditionally termed "masculine" and "feminine"):

(3) a. **y** gath (*cath*)
the cat.F
'the cat'

b. y cathod
the cat.PL
'the cats'

c. y carlwm
the stoat.M
'the stoat'

d. y carlymod
the stoat.M.PL
'the stoats'

A second instance of a morphosyntactic restriction is shown in (4), this time involving a restriction on the trigger. Feminine singular nouns trigger soft mutation on each following adjective (4a), but masculine and plural nouns do not trigger mutation on following adjectives (4b):

(4) a. y gath fawr ddu (*mawr, du*)
the cat.F big black
'the big black cat'

b. y cathod mawr
the cats big
'the big cats'

The mutation triggered by feminine singular nouns also applies if one of a string of adjectives has a non-mutable initial consonant, as in *y gath ifanc ddu* 'the young black cat', where the adjective *ifanc* 'young' does not have a mutable initial but the following adjective does, so bears soft mutation (*du* > *ddu*). Various approaches to handling adjective mutation are discussed in Borsley et al. (2007). As with all other mutations in Welsh, the mutation is strictly local, and the trigger must immediately precede and c-command the target. A plausible structure for the phrase in (5) is shown in Figure 1:

(5) gardd fawr, breifat (*mawr*, *preifat*)
garden.F big private
'a large, private garden'

In Figure 1, the assumption is that a feminine singular NP triggers soft mutation on its sister, and the adjectives are each adjoined to form a new NP which in turn is a trigger for the mutation.

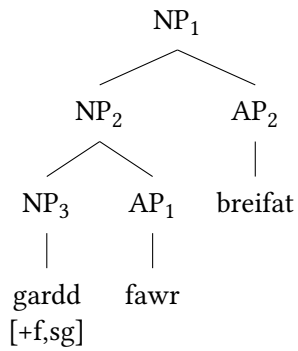


Figure 1: Tree for (5)

So, the trigger for *mawr* > *fawr* is the feminine singular NP_3 *gardd*, and the trigger for *preifat* > *breifat* is the feminine singular NP_2 *gardd fawr*, which has inherited its number and gender features from *gardd*.

Note that the mutation in examples like (5) is not plausibly treated as morphological agreement of an adjective with a feminine singular head noun. In (6), the AP undergoes soft mutation as expected following the feminine singular noun, with the mutation realized as usual on its initial segment, giving *tra* > *dra*. But the head adjective, *campus* 'splendid', does not agree with the feminine singular head N to give soft mutation; instead, it undergoes aspirate mutation (AM), triggered by the modifier *tra*:

(6) gardd [$_{AP}$ dra^{A} champus] (*tra, campus*)
garden.F quite splendid(+AM)
'a quite splendid garden'

Example (6) also illustrates nicely the strict locality of mutation in Welsh: as indicated, mutation is always triggered by an immediately preceding c-commanding item, and the most local mutation trigger always takes precedence.

2.3.2 Marked word orders

Adjectives in Welsh are generally post-nominal, but some adjectives occur, or may occur, pre-nominally. The following element, generally the head noun, then undergoes soft mutation irrespective of number and gender:

(7) a. hen gathod (*cathod*)
old cat.PL
'old cats'

b. hen garlymod (*carlymod*)
old stoat.PL
'old stoats'

Again, the triggering adjective immediately precedes and, under reasonable assumptions, c-commands the target (see also Willis 2006). Tallerman (1999) treats this as one instance of a more general triggering environment for soft mutation in which a marked word order is signaled by mutation occurring on the head: in Welsh, head-initial word order is unmarked, while head-final order is marked[1]:

(8) The markedness hypothesis:
In any phrase XP with an unmarked word order [X^0 α], X^0 bears soft mutation if it is *preceded* by α.

While Adj-N marked word order is relatively easy to find, we also see the effects of this mutation trigger in far less common environments, especially in Literary Welsh. For instance, numerals in Welsh are standardly pre-nominal, as in (9a), but in the marked N-Num word order in (9b), soft mutation is triggered on the numeral:

[1]There is no support in Welsh for the proposal that NPs should be analysed as Determiner Phrases (see, for instance, Borsley et al. (2007): ch. 3.1; 5.6). The head of phrases consisting of D+N and Q+N will therefore be assumed to be N in what follows.

(9) a. tri dyn (Num-N unmarked)
three man
'three men'

b. dynion dri (*tri*) (N-Num marked, SM on Num)
men three
'three men'

Similarly, quantifiers are standardly pre-nominal, as in (10a); unlike adjectives, these do not trigger soft mutation on the following noun, so for instance *llawer* 'many', *pob* 'every' and *peth* 'a bit of' are all followed by the radical initial consonant. However, in the marked word order, (10b), the post-nominal quantifier undergoes soft mutation:

(10) a. llawer dyn (Q-N unmarked)
many man
'many men'

b. dynion lawer (*llawer*) (N-Q marked, SM on Q)
men many
'many a man'

The assumption here is that the noun is the head of its phrase, rather than the numeral or the quantifier.

However, it is not so clear whether the markedness hypothesis handles the situation in (11), which has a string of pre-nominal adjectives. Here we see that the head noun undergoes soft mutation, as predicted by the hypothesis, but each adjective following the first in the series is also a target for soft mutation, triggered by the preceding adjective:

(11) yr hen brif breswylfa (*prif, preswylfa*)
the old main residence
'the former main residence'

In this example, the order of the adjectives (*hen* > *prif*) gives rise to the meaning 'former' for *hen*, whereas if the meaning is the standard 'old', the order is reversed:

(12) y prif hen gymeriadau (*cymeriadau*)
the main old character.PL
'the main old characters'

A reasonable assumption is that (11) does in fact have a marked word order, since *hen* has the atypical 'former' meaning here, and therefore appears in an atypical position compared to examples like (12). Note also that equative, comparative and superlative adjectives, which at least in standard Welsh have an unmarked pre-nominal position, do not generally trigger soft mutation.

However, the mutation of the adjective *prif* > *brif* (in 11) remains unaccounted for by the formulation of the markedness hypothesis in (8), since this only applies to heads and their modifiers, and not to the case in (11), where a modifier triggers soft mutation on another modifier.

2.3.3 Syntactic soft mutation

Finally in this section I briefly outline what we can call syntactic soft mutation – mutation which is triggered in a specific syntactic context in Welsh, unique amongst the Celtic languages. We will see that phrases (XPs) are triggers for soft mutation and trigger the mutation on a following complement. The "XP trigger hypothesis" was proposed by Borsley & Tallerman (1996) and further defended by Borsley (1999), Tallerman (2006), Tallerman (2009), and Borsley et al. (2007):

(13) XP trigger hypothesis (Borsley 1999):
A complement bears soft mutation if it is immediately preceded by a phrasal sister.

Thus, the trigger for soft mutation both c-commands and immediately precedes the target. The formulation of the XP trigger hypothesis in (13) accounts for numerous contexts for soft mutation in Welsh which have traditionally been regarded as distinct environments. In (14a) we see the context sometimes known as "direct object mutation", illustrating the mutation of the object of a finite verb. Here, the XP trigger is the subject NP (in bold), *y ddynes*, and the direct object which is the target for the mutation is a complement of the finite verb: *beic* > *feic*:

(14) a. Prynodd **y ddynes** feic. (*beic*)
buy.PAST.3S the woman bike
'The woman bought a bike.'

b. Beic / *feic brynodd y ddynes _____.
bike / bike(+SM) buy.PAST.3S the woman
'The woman bought *a bike*.'

The XP trigger hypothesis correctly predicts the mutation in (14a) under the assumptions of a surface-orientated phrase structure in which the subject is the

sister of the direct object; see for instance Borsley (1999). However, in (14b) where the direct object is fronted, it is nonetheless a complement of the verb but does not follow an XP trigger and therefore does not mutate; nor does a sentence fragment object mutate, for the same reason. Such data show that the idea that 'direct object mutation' reflects case-marking, as proposed by Roberts (2005), is untenable. There is also no soft mutation on the direct object of a non-finite verb, which is again predictable under the XP trigger hypothesis since there is no preceding XP trigger:

(15) Mae 'r ddynes wedi prynu beic /*feic.
be.PRES.3S the woman PERF buy.INF bike /bike(+SM)
'The woman has bought a bike.'

There is, though, soft mutation on the VP in (16), realized as usual on the initial element of the phrase, in this case the non-finite verb: *gwerthu* > *werthu*. Here, the trigger is again the subject NP. The XP trigger hypothesis predicts this mutation on the reasonable assumption that the non-finite VP is a complement of the finite auxiliary *gwnaeth* 'did':

(16) Gwnaeth **y ddynes** [$_{VP}$ werthu beic]. (*gwerthu*)
do.PAST.3S the woman sell.INF bike
'The woman sold a bike.'

Note that again, there is no soft mutation on the direct object (*beic*) in (16), since no XP (or other trigger) precedes it. Further arguments against the "mutation as case" hypothesis are outlined in Tallerman (2006).

The XP trigger hypothesis correctly predicts the occurrence of soft mutation in various other contexts. Here, I illustrate just two. The first is traditionally known as 'soft mutation following a parenthesis'; compare (17a) and (17b):

(17) a. Mae ci mawr yn yr ardd.
be.PRES.3S dog big in the garden
'There's a big dog in the garden.'

b. Mae **[$_{PP}$ yn yr ardd]** gi mawr. (*ci*)
be.PRES.3S in the garden dog big
'There's a big dog in the garden.'

The usual constituent order is shown in (17a), while (17b) illustrates a less common word order, with the PP occurring in a focalized position before the subject. In (17a) there is no XP preceding the subject *ci mawr* 'a big dog', while in (17b),

the PP precedes the subject, triggering soft mutation, which is realized as usual on the initial element of the target: *ci* > *gi*. On the assumption that subjects are also complements of the finite verb or auxiliary (Borsley 1999), the PP is correctly predicted by the XP trigger hypothesis to be the phrasal trigger for the mutation here. Such 'parentheses' can be PPs or some other intervening constituent, such as an Adverb Phrase, and these occur in various syntactic contexts. For instance, in (18), where the non-finite VP *dysgu'r gelfyddyd* 'to learn the art' is a complement of the verb *ffaelio* 'fail', an intervening adverbial occurs before the VP, triggering soft mutation on its initial element, just as predicted by the XP trigger hypothesis (Tallerman 2006):

(18) ... yn ffaelio [$_{\text{AdvP}}$ **'n glir lân]** [$_{\text{VP}}$ ddysgu 'r gelfyddyd]
PROG fail.INF PRED completely learn.INF the art
'completely failing to learn the art' (adapted from Morgan 1952: 432)

Without the Adverb Phrase trigger, the VP would not bear soft mutation.

A further illustration of the power of the XP trigger hypothesis is given in (19):

(19) Dw **i** [$_{\text{AP}}$ lawn mor grac â chi]. (*llawn*)
be.PRES.1S I full as angry as you
'I'm just as angry as you.'

Here, an AP predicate complement bears soft mutation (*llawn* > *lawn*), triggered by the subject NP *i*. Other non-nominal constituents, such as PPs, can also occur as predicate complements, and as predicted these also bear soft mutation (Tallerman 2006).

Finally, note that it is crucial in (13) that the XP trigger hypothesis refers to complements, and not just any constituent, as the target for soft mutation triggered by a preceding XP. In (20), we have a direct object which consists of a set of conjoined NPs:

(20) Prynon **nhw** [fara, menyn, a chaws]. (*bara*)
buy.PAST.3P they bread butter and cheese
'They bought bread, butter and cheese.'

Here, it is the whole direct object *bara, menyn, a chaws* that forms the complement of the verb. This follows a trigger for soft mutation, the subject *nhw*, and the direct object therefore bears soft mutation on its initial element as usual: *bara* > *fara*. Crucially, each individual conjunct is not a complement, so no mutation is triggered, for instance, by the first conjunct onto the second: **fenyn*. And the

final conjunct takes aspirate mutation, triggered by the immediately preceding trigger *a* 'and'. As usual, if the initial element of the first conjunct happened not to have a mutable initial consonant, then the direct object would show no other signs of soft mutation.

Thus, we see that the XP trigger hypothesis correctly predicts the occurrence of soft mutation in a wide variety of syntactic contexts, more of which are discussed and illustrated in the references shown. We have also seen that targets for syntactic soft mutation include objects, subjects, verb phrases headed by non-finite verbs, and predicates in copular clauses.

3 Further complexities of Welsh mutation

3.1 Not all mutations have a trigger

Though most mutations are triggered, some are not. The most common context involves adverbials of time, which at least in standard Welsh bear soft mutation in any sentence position, realized as usual on the initial element wherever this has a mutable initial:

(21) Ddwy flynedd yn ôl, diflannodd hi. (*dwy*)
two year ago disappear.PAST.3S she
'Two years ago, she disappeared.'

(22) Fis yn ôl dechreuodd y broblem. *(mis)*
month ago begin.PAST.3S the problem
'A month ago the problem began.'

Probably the most frequently occurring example is *ddoe,* 'yesterday', which is never **doe.*[2]

Other adjuncts also often bear soft mutation optionally, such as the measure adverbial here:

(23) 'Dan ni 'n byw dair milltir o 'r dre'. *(tair)*
be.PRES.1P we PROG live.INF three mile from the town
'We live three miles from town.'

[2]There is generally no soft mutation on *bore* 'morning' or *prynhawn* 'evening', whatever the position. Idiosyncratic exceptions to mutation in Welsh are not at all uncommon.

Since adjuncts are not complements, such examples do not fall under the XP trigger hypothesis. The mutation of adjuncts is subject to dialectal and idiolectal variation, but where it does occur, it consistently lacks any trigger.

Various other items always appear in their soft mutation form, such as the negative *fawr* (< *mawr*, 'big'), which means 'not much':

(24) Does fawr yn newid.
NEG.be.PRES.3S not.much PROG change.INF
'Not much changes.'

Fawr also occurs in expressions like *fawr o neb* 'not many people' (*neb* 'no one'), and *fawr o ddim* 'not much' (literally, 'not much of nothing'):

(25) Be' wnest ti? Fawr o ddim.
what do.PAST.2S you not.much of nothing
'What did you do? Not much.'

Note that the occurrence of soft mutation in (25) contrasts with the situation seen in sentence fragment objects, which do not bear soft mutation, since they do not have an appropriate XP trigger. It is perhaps the case that the negative *fawr* is more properly regarded as a lexicalized form, as its mutation is not triggered. Other lexicalized soft mutation forms occur, such as *gartre(f)* 'at home, homewards' (cf. *cartref* 'home') and *bellach* 'any more' (cf. *pell* 'far'); these have fixed soft mutation and only occur in that form (Morgan 1952: 257ff).

3.2 Not all mutation triggers are overtly realized

We showed above that the singular proclitics which appear on nouns and non-finite verbs in agreement contexts are triggers for various mutations; see (1). However, in the colloquial language the proclitics are often omitted. In this case, the mutation that the proclitic would have triggered may remain, as these examples show (NM is nasal mutation). The relevant proclitics are shown in parentheses:

(26) Dw i wedi (**ei**) weld o o'r blaen. (*gweld*)
be.PRES.1S I PERF 3MS see.INF(+SM) him before
'I've seen him before.'

(27) Hwyrach dy fod ti wedi (**yn/fy**) ngweld i. (*gweld*)
perhaps 2S be.INF you PERF 1S see.INF(+NM) me
'Perhaps you've seen me.'

In the most colloquial register (cf. Borsley et al. 2007: Sections 4.1.5; 4.2.2), both the proclitic and the mutation are absent, but the forms shown above are also acceptable in spoken Welsh. Crucially, even if the proclitic that would be a mutation trigger is phonetically unrealized, it continues to block any mutation from an immediately preceding overt trigger. For instance, (28) contains the aspect marker *heb* (lit. 'without'), a trigger for soft mutation:

(28) Dw i heb (ei) gweld/*weld hi ers misoedd
be.PRES.1S I without (3FS) see.INF/*see(+SM) her since months
'I haven't seen her in months.'

Here, more formal Welsh would have an overt agreement proclitic, *ei* (3FS), which triggers aspirate mutation. Since *gweld* 'see' does not have a voiceless initial stop it cannot take this mutation, so a formal variant of (28) would have *ei gweld hi.* If *ei* is unrealized, *gweld* appears to immediately follow the trigger for soft mutation, *heb*. But the unrealized proclitic blocks the soft mutation, as (28) shows: **weld*; even if phonetically unrealized, the proclitic continues to occupy a syntactic position as the closest mutation trigger.

In fact, numerous small functional elements are phonetically unrealized in the spoken language, yet the mutations they trigger can remain. The clause-initial complementizers (often termed particles) are typical examples. For instance, affirmative root clauses can have an optional complementizer *mi* or *fe*, both triggers for soft mutation, and the mutations may remain even when the triggers are not overtly present. Similarly, many place names start, at least historically, with the definite article *y(r),* which is a trigger for soft mutation on a following feminine singular noun, as in (3) above. The article is often unrealized in modern contexts, but the mutation remains on the initial element of feminine singular placenames: *Drenewydd* < *Y Drenewydd* (*tre' > dre'* 'town'), 'Newtown'; *Gaerwen* < *Y Gaerwen* (*caer > gaer* 'fort').

Finally, I turn to non-overt mutation triggers under the XP trigger hypothesis. Within the generative tradition, the two 'Case-marked' empty category NPs *pro* and *wh*-trace are recognized as placeholders for the overt phrases that they replace. As predicted by the XP trigger hypothesis, these empty categories continue to trigger mutation, and indeed to block mutation, just as if they were overt. The first example shows the empty category *pro,* the null subject of a finite clause which occurs in more formal varieties of Welsh:

(29) Prynodd *pro* delyn. (*telyn*)
buy.PAST.3S harp
'He/she bought a harp.'

The direct object, *telyn*, bears soft mutation under the XP trigger hypothesis just as it would with an overt subject pronoun or full NP trigger. And in (30) we see that *wh*-trace in the subject position of a finite clause also triggers soft mutation under the XP trigger hypothesis:

(30) y ddynes brynodd *wh-t* delyn (*telyn*)
the woman buy.PAST.3S harp
'the woman who bought a harp'

See Tallerman (2006) and Tallerman (2009) for a full account of the role played by empty categories in mutation; further discussion is found in Borsley et al. (2007): ch. 7.

3.3 Not all lexical items bear mutation

It is worth briefly mentioning that not all words undergo mutation, even if they are found in a triggering environment. This set of non-mutating words is relatively small, and generally consists of functional or semi-functional elements: some never mutate, such as *dy*, the second person singular proclitic (i.e., 'your'), and *tra* 'while'; other items do not standardly mutate or may vary between retaining the radical and bearing mutation. The adjectival modifiers *mor* and *cyn* 'so, as', for instance, do not mutate (nor does the preposition *cyn* 'before'), though for some speakers the pre-adjectival *mwy* 'more' does take soft mutation (> *fwy*). A few content words fail to mutate; for instance, the very old loanword *braf* 'fine, splendid' generally does not undergo soft mutation, with the apparent exception of some southern dialects. Various prepositions typically occur in a lexicalized soft mutation form and cannot take further mutation, e.g. *dan* 'under', *dros* 'across, over' and *drwy* 'through'. Etymologically these derive from forms with radical initial *t-*, and this still occurs in certain set expressions, such as *tros y pen a'r clustiau* 'head over heels' (lit. 'over the head and the ears').

This section has briefly presented a few of the further complexities of Welsh mutation, though many more exist. More details concerning the variation in mutations, the inability of various items to mutate and the occurrence of lexicalized mutations can be found in Morgan (1952).

4 Some idiosyncrasies of mutation

In this section, I explore some of the less obvious instances of mutation in Welsh. We will see that there are variations in the perceived mutability of forms, and

in the effects of mutation. Unless otherwise stated, the examples of mutation in this section are soft mutation.

4.1 Mutation and variation in the radical initial segment

First, though we saw in Section 1 that the standard consonantal inventory for mutation consists of only nine consonants, in fact a tenth consonant is typically added, especially by northern speakers, in the form of a loan consonant (Griffen 1974). The voiceless affricate [ʧ] does not occur in the standard set of mutable consonants[3], but appears in initial position in a few well-established loanwords, such as *tsiaen* 'chain' and *tsiocled* 'chocolate'. In a mutation context such as (31), where the conjunction *neu* 'or' is a soft mutation trigger, the noun undergoes mutation to give an initial voiced affricate [ʤ]:

(31) (Peidiwch byth â chysylltu) rhaff **neu** dsiaen... (*tsiaen)*
(Never connect) rope or chain
'Never connect a rope or a chain...'

Similarly, the loanword *tsips* 'chips' with initial [ʧ] is common in spoken Welsh and mutates to *jips* with initial voiced [ʤ] in soft mutation contexts (here, an XP trigger hypothesis context):

(32) Gymi **di** jips? (*tsips*)
take.FUT.2S you chips
'Do you want some chips?'

The addition of the affricate [ʧ] to the set of mutable consonants was noted as far back as the start of the last century by Fynes-Clinton (1913), who cites examples such as (33), where loaned English 'chap', *tshiap*, with a voiceless initial affricate, mutates in a soft mutation context (following a pre-nominal adjective) to give a voiced initial affricate. Note that there is no standard way to spell words with initial affricates, so a variety of solutions exist, as these examples show:

(33) yr hen [ʤap]
the old chap
'the old chap'

[3]The affricates, both voiceless and voiced, are not part of the traditional consonant inventory at all (Hannahs 2013) but have developed due to palatalization in native words such as *diawl,* 'devil', with initial [ʤ].

For some speakers, as a reviewer notes, [ʧ] can also undergo nasal mutation following a trigger such as *fy/yn* 'my', giving examples such as *tsiocled* > *fy/yn nhiocled* 'my chocolate'.

Second, we turn to instances of speaker variation over the choice of 'correct' initial radical consonant, apparently at least partly motivated by the soft mutation form of the word. In fact, even in the standard language doublets occur, so that alternative citation forms of some common words exist:

(34) a. benthyg / menthyg 'borrow' (V)
b. benyw / menyw 'woman' (N)
c. bigwrn / migwrn 'ankle, knuckle' (N)

It seems reasonable to suggest that such doublets occur with initial *b-* or *m-* because both have the soft mutation form [v], leading to two possible choices of radical form. Since *benyw/menyw* is also a feminine noun, it frequently occurs in a soft mutation context following the definite article: *y fenyw*; *bigwrn/migwrn*, however, is masculine.

Loanwords from English are especially susceptible to reanalysis of the initial consonant. If the loanword has initial [v] in English, there is a tendency for this to be perceived as the soft mutation form of the word: only a handful of native Welsh words have [v]-initial citation forms, and these comprise either small function words such as *fy* 'my' or grammaticalized items such as *faint* (< *pa faint*) 'how much, how many' (Welsh <f> = [v]). So, in southern dialects, instead of standard *fecso, fecsio* (< 'vex') we find the form *becso*, where a perceived soft mutation form is reanalysed to have an initial consonant expected in native words, [b]:

(35) fecso / becso 'worry' (V)

Other doublets or triplets are frequently attested in English loanwords with initial [v]:

(36) a. '(ad)vantage' > *mantais*, but also *fantais*
b. 'venture' > *mentro*, but also *fentro*
c. 'varnish' > *barnais*, but also *marnais, farnais*
d. 'vicar' > *bicar*, but also *ficer*

In all these examples, initial [v] is apparently perceived to represent a soft mutation form, just as it would in native content words; it is therefore reanalysed to have the 'correct' initial consonant in the radical form.

Speaker variation concerning the radical form of a word also occurs with words that could be perceived as either [g]-initial or vowel-initial in the radical form. Since [g] deletes under soft mutation, speakers may reanalyse vowel-initial words with a perceived 'missing' radical initial consonant. Again, some of this variation occurs in standard Welsh, giving rise to doublets such as *allt/gallt* 'slope, hill'; *ordd/gordd* 'hammer, mallet', *ogof/gogof* 'cave'. These nouns are feminine, and clearly often occur following the definite article, which, as we saw in Section 1, is a trigger for soft mutation; this might account for some variation. But *wyneb/gwyneb* 'face' is masculine, so would not lose the initial [g] following the article. In all such cases, the vowel-initial form of the noun seems standardly to be the etymon, while the [g]-initial form is the reanalysed form. Conversely, the standard citation form of the loanword from English 'honest' is *gonest*, though *onest* also occurs as an alternative citation form. Again, motivations for this second reanalysis are not hard to find, since adjectives often occur in a triggering context for soft mutation. For instance, adjectives in the predicative position follow the predicate marker *yn*, a trigger for soft mutation:

(37) Mae 'r dyn **yn** onest. (*gonest*)
be.PRES.3S the man PRED honest
'The man is honest.'

In spoken Welsh, the verb *addo* 'promise' is reanalysed as (non-standard) *gaddo* by some speakers, reconstructing a 'missing' radical initial [g]. This inter-speaker variation occurs quite naturally in exchanges such as (38):

(38) a. Speaker A: Addo?
promise.INF
Speaker A: '(Do you) promise?'
b. Speaker B: Gaddo!
promise.INF
Speaker B: '(I) promise!'

Finally in this section, we also find dialectal variation over the choice of radical initial consonant, especially in loanwords (Morgan 1952:455f):

(39) a. *tamp* > SM form *damp* (northern form) 'damp'
b. *damp* > SM form *ddamp* (southern form) 'damp'

Here we see that if *damp* is perceived as the soft mutation form, then it must be reconstructed to have the radical form *tamp*.[4]

[4]A reviewer for this volume suggests that the variation may be due to historical differences in

4.2 Extending mutation to nonce loans

We have already seen some of the effects of mutation in loanwords in previous sections, and mutation may either recreate the phonology of the loaning language or destroy it. Examples of the former include *mantais* > *fantais* and *gonest* > *onest*; the latter includes examples such as *tsips* > *jips*.

Instances of mutation in established loanwords are commonplace; Welsh has thousands of such loans, and many of course begin with a mutable initial consonant. Some typical examples of established loans are shown here, all following triggers for soft mutation:

(40) a. *cwpl o fois* 'couple of boys' (*bois*)
b. *bocs llawn o dŵls* 'a box full of tools' (*tŵls*)
c. *prif falerina* 'principal ballerina' (*balerina*)
d. gofyn i gwsmeriaid i foicotio 'r archfarchnad (*boicotio*)
ask the customers to boycott the supermarket
'Ask the customers to boycott the supermarket.'
e. *dy basport di* 'your passport' (*pasport*)
f. *yn y farced* 'in the market' (*marced*)
g. *bach o farsipán* 'a bit of marzipan' (*marsipán*)
h. *bach o dreiffl* 'a bit of trifle' (*treiffl*)

Loans are also incorporated into the standard system for the other mutation series, as (41) shows with aspirate mutation following the feminine singular proclitic *ei*:

(41) ei thedis hi (*tedis*)
3FS teddies(+AM) her
'her teddies'

Not only well-established loans, but also nonce loans are subject to mutation, entirely regularly. Some typical examples with soft mutation are shown in (42); single quotes indicate a clearly English word and pronunciation:

(42) a. *dy [ð]êt di* 'your date' (*dêt* – cf. the established loan *dât*, meaning a date on the calendar)

voice onset time between English and different dialects of Welsh, with north Walians hearing English /d/ as voiceless.

b. *ym [mh]rofile i* 'my profile' (English pronunciation; cf. the standard loan *proffil*, with a distinct pronunciation)
c. *ti'n [v]echanig* 'you're a mechanic' (*mechanig*, not the standard loan *mecanic*)
d. *rhyw [ð]ampener* 'kind of dampener' (*dampener)*
e. *cwpl o [ð]rincs* 'couple of drinks' (*drincs*)
f. *siarad efo ei [g]ontacts* 'speak to his contacts' (*contacts*)
g. *mwy o [v]uscles* 'more muscles'
h. *dy [v]acon butty* 'your bacon butty'
i. *unrhyw '[v]oiler'* 'any boiler'

In (42a), the English word for a date with a boyfriend or girlfriend is loaned, and bears soft mutation following the second person proclitic *dy*. In (42b), there exists a standard loanword *proffil*, with standard Welsh phonology, but in this (oral) example the English word *profile* was used, and the nonce loan bears nasal mutation following the first person proclitic *yn* here assimilated to *ym*. In (42h), it is entirely clear that *bacon butty* is a nonce loan, since it does not follow Welsh head-initial word order, cf. *brechdan cig moch* (*brechdan* 'sandwich', *cig moch* 'bacon') 'bacon sandwich'.

Clearly, then, mutation is entirely productive in Welsh, and both nonce loans and established loans are integrated fully into the mutation system.

4.3 Mutation of non-citation and novel citation forms of words

Given the fluidity of mutation illustrated in previous sections, it is not too surprising that a variety of non-standard instances of mutation occur quite commonly. These include what are sometimes probably just nonce slips, though these are often not merely speech errors, but also occur in written Welsh, even of a fairly formal variety. This section illustrates with various instances of soft mutation.

First, the deletion of initial syllables in Welsh is quite common, even if these are penultimate syllables and therefore stressed, the regular stress being on the penult. In southern varieties, for instance, we find *hefyd* > *'fyd* 'also' and *eto* > *'to* 'again, yet'. In some cases, the reduction gives rise to a new or alternative citation form, which may then produce a mutable initial consonant distinct from that of the full form:

(43) *baned* (< *paned*) from *cwpanaid* 'cupful, cuppa'

In this case, the standard citation form for 'cupful, cuppa' is *cwpanaid*, a noun which has a mutable initial consonant, [k], but which is masculine, so does not bear soft mutation following the definite article: *y cwpanaid*. However, the reduced form *paned, panad* is the normal form in spoken Welsh, and somewhat surprisingly, this is a feminine noun, so soft mutation is triggered after the definite article.[5]

Similarly, the full form of the noun meaning 'bird' is *aderyn*, but the common spoken form is *deryn*. This noun is masculine, so does not bear soft mutation following the definite article. However, in other mutation environments this reduced form is always treated as a normal citation form, and so mutates in appropriate contexts. Here the soft mutation occurs following an XP trigger:

(44) a. Mi welais **i** dderyn bach. *(deryn)*
PRT see.PAST.1S I bird small
'I saw a small bird.'

b. Mae **yna** dderyn bach. (*deryn*)
be.PRES.3S there bird small
'There's a small bird.'

Note that the plural form of the noun is vowel-initial, *adar* 'birds', so the formal citation form of the singular noun is presumably quite salient.

The full form of the noun meaning 'Christmas' is *Nadolig*, often contracted to *Dolig*, which then undergoes soft mutation following a trigger such as predicative *yn*:

(45) Mae 'n Ddolig. (*Dolig*)
be.PRES.3S PRED Christmas
'It's Christmas.'

Second, double mutations are also not uncommon in speech and even in writing: a form of a word that is the soft mutation form is treated as the radical, and then undergoes further mutation. Here, the trigger for soft mutation is the predicate marker *yn* (reduced to *'n* after a vowel). The citation form is *parod,* soft mutation form *barod*; if *barod* is then perceived as the radical form, it mutates to become **farod*:

[5]The root noun *cwpan* 'cup' can be masculine or feminine, and a reviewer suggests that *cwpanaid* may have originated in a dialect in which it is feminine.

(46) Dw i 'n farod i helpu. (✓*parod* > ✓*barod* > **farod*)
be.PRES.1S I PRED ready to help radical (+SM) (SM of SM form)
'I'm prepared to help.'

(47) ...pan fydd o 'n farod i 'r twrnamaint i ddechrau
when be.FUT.3S he PRED ready for the tournament to begin
'...when he is ready for the tournament to begin'

Similarly, the citation form of the verb for 'vote' is *pleidleisio*, SM form *bleidleisio*, but the doubly mutated form **fleidleisio,* following a soft mutation trigger, is not hard to find in internet searches:

(48) Ar ôl y rhyfel gafodd fenywod dros 30 yr hawl i fleidleisio.
after the war get.PAST.3S women over 30 the right to vote.INF
'After the war women over 30 got the right to vote.'

(49) y rhesymau dros fleidleisio cynnar
the reasons for vote.INF early
'the reasons for early voting'

We thus see the productivity of the Welsh mutation system, its extensions to novel and nonce loans, and the existence of both idiolectal and dialectal variation in mutation and in the citation forms of words.

5 A final conundrum

I now briefly present an interesting conundrum involving the definite article and the soft mutation which it triggers on a following feminine singular noun. The definite article has three surface forms, *'r* [r], *yr* [ər] and *y* [ə]. The enclitic form *'r* occurs after a vowel-final word and takes precedence over the other two forms; it will not be further discussed here. The variant *y* occurs before a consonant, as in *y llyfr* 'the book' and *yr* occurs elsewhere, i.e., before vowels, e.g. *yr afon* 'the river', and before glides and [h]. Now consider the following data:

(50) a. gardd [FEM. SG.]
'garden'

b. yr ardd [g] > ∅
the garden(+SM)
'the garden'

In (50b) we see mutation on the noun following the article, resulting in the soft mutation form: *gardd* > *ardd*. Since the noun is now vowel-initial, the article must be *yr*. So *yr* occurs not only before nouns which are canonically vowel-initial (*yr afon*), but also before nouns in which the initial vowel results from mutation: *yr ardd*. The soft mutation form of the noun determines the form of the article, yet the article triggers the mutation. Here, then, we have a paradox: the mutation (*gardd* > *ardd*) is triggered by the preceding article, but the correct form of the article (*y / yr*) cannot be chosen unless the initial segment of the noun is known – which in turn depends on whether or not mutation has occurred.

The solution presented by Hannahs & Tallerman (2006) relies on the notion of staggered lexical insertion (Emonds 2000, 2002), in which content words are inserted into a derivation before functional elements.

(51) $[\_\_]_{\text{Def Art}}$ $[\text{gardd}]_{\text{N}} \Rightarrow [\_\_]_{\text{Def Art}}$ $[\ \text{ardd}]_{\text{N}}$

Here, the soft mutation on the noun is triggered by the as-yet-unfilled slot for the definite article; the article itself is inserted later in the derivation, in its correct form *yr*, since it can now 'see' the initial segment of the noun.[6]

Independent support for this analysis comes from the fact that various small functional elements are often unrealized; see (26) and (27). But the mutations triggered by these phonetically null items remain, at least in some registers; therefore, unfilled slots that trigger various mutations are independently needed. Moreover, these unfilled slots have a syntactic presence in that they block an overt preceding item from triggering some mutation, as we also saw in (28). Example (52) shows the full form of a PP with an overt agreement proclitic *ein*:

(52) yn ein tŷ ni (*ein* is not a mutation trigger)
in 1P house us
'in our house'

It is common in speech for the proclitic to be unrealized, as in (53a). The noun *tŷ* then appears to follow a trigger for nasal mutation, *yn*, but this does not trigger any mutation here, as shown in (53b), because the unfilled slot for the proclitic blocks the mutation:

(53) a. yn ~~ein~~ tŷ ni
in ~~1P~~ house us
'in our house'

[6] As a variant of this analysis, a reviewer suggests that the input to the definite article is *y(r)* in all cases, where the [r] is present, but has no phonological slot. If the following word begins with a vowel, the [r] is saved. If not, it is deleted.

b. * yn nhŷ ni
in house(+NM) us
'in our house'

c. yn nhŷ Alun
in house(+NM) Alun
'in Alun's house'

By contrast, in (53c) where there is no proclitic (as these only occur with pronominal possessors), the preposition *yn* does precede the noun, and this takes NM: *tŷ* > *nhŷ*.

6 Future directions

The mutation system is clearly still active in the modern Welsh language, as previous sections have shown. The vast majority of environments for mutation are contexts for soft mutation, and the other two series are very limited, each having only a few triggers in the form of small functional elements. However, one relatively recent development not yet mentioned is the strong tendency for soft mutation to occur in certain non-traditional contexts, spreading to environments which are formally contexts for AM or NM. For instance, *yn* 'in' is a trigger for nasal mutation in standard Welsh, but is often followed instead by soft mutation. Thus, we might find *yn Fangor* 'in Bangor' instead of the standard *ym Mangor*. More details concerning sociolinguistic aspects of the modern mutation system can be found in Webb-Davies (forthcoming). The full picture of dialectal and idiolectal variation in the mutation system remains to be investigated, however. In addition, changes are in progress owing to the numbers of new speakers, typically those who do not come from a Welsh-speaking household but have attended a Welsh-medium school.

This paper has outlined some of the more interesting features of Welsh mutation, yet numerous aspects of the mutation system are only touched on here. The 'bible' of Welsh mutations is Morgan (1952), though the fact that this is in Welsh precludes a wide readership. A reasonably comprehensive list of mutation triggers can be found in many pedagogical grammars; perhaps the most accessible of these is King (2015), which also contains useful details of dialectal variation. Discussion and formal analysis of various aspects of mutation is found in Borsley et al. (2007), and discussion of the phonological aspects of mutation is found in Hannahs (2013).

Acknowledgements

I am grateful to two anonymous reviewers for suggesting very helpful changes and corrections, and to Bob Borsley and S.J. Hannahs for their comments on an earlier version of this paper. All remaining errors are my own.

References

Borsley, Robert D. 1999. Mutation and constituent structure in Welsh. *Lingua* 109. 263–300. DOI: 10.1016/S0024-3841(99)00019-4.

Borsley, Robert D. & Maggie Tallerman. 1996. Phrases and soft mutation in Welsh. *Journal of Celtic Linguistics* 5. 1–49.

Borsley, Robert D., Maggie Tallerman & David Willis. 2007. *The syntax of Welsh*. Cambridge: Cambridge University Press.

Emonds, Joseph E. 2000. *Lexicon and grammar: The English syntacticon*. Berlin: Mouton de Gruyter.

Emonds, Joseph E. 2002. A common basis for syntax and morphology: Tri-level lexical insertion. In Paul Boucher (ed.), *Many morphologies*, 266–290. Somerville, MA: Cascadilla Press.

Fynes-Clinton, O.M. 1913. *The Welsh vocabulary of the Bangor district*. Oxford: Oxford University Press.

Griffen, T.D. 1974. The development of Welsh affricates: A change through borrowing. *Lingua* 34. 149–165. DOI: 10.1016/0024-3841(74)90014-X.

Hammond, Michael. 2019. Voiceless nasals in Welsh. *Journal of Celtic Linguistics* 20. 31–60. DOI: 10.16922/jcl.20.3.

Hannahs, S.J. 2013. *The phonology of Welsh*. Oxford: Oxford University Press.

Hannahs, S.J. & Maggie Tallerman. 2006. At the interface: Selection of the Welsh definite article. *Linguistics* 44. 781–816. DOI: 10.1515/LING.2006.025.

King, Gareth. 2015. *Modern Welsh: A comprehensive grammar*. 3rd edn. London: Routledge.

Morgan, T.J. 1952. *Y treigladau a'u cystrawen*. Cardiff: Gwasg Prifysgol Cymru.

Roberts, Ian. 2005. *Principles and parameters in a VSO language: A case study in Welsh*. Oxford: Oxford University Press.

Tallerman, Maggie. 1999. Welsh soft mutation and marked word order. In Michael Darnell, Edith Moravcsik, Frederick Newmeyer, Michael Noonan & Kathleen Wheatley (eds.), *Functionalism and formalism in linguistics*, vol. 2, 277–294. Amsterdam: John Benjamins. DOI: 10.1075/slcs.42.15tal.

Tallerman, Maggie. 2006. The syntax of Welsh "direct object mutation" revisited. *Lingua* 116. 1750–1776. DOI: 10.1016/j.lingua.2004.10.004.

Tallerman, Maggie. 2009. Phrase structure vs. dependency: The analysis of Welsh syntactic soft mutation. *Journal of Linguistics* 45. 167–201. DOI: 10 . 1017 / S0022226708005550.

Webb-Davies, Peredur. Forthcoming. Initial consonant mutation in Welsh. In Joseph Eska, Paul Russell, Peadar Ó Muircheartaigh & Silva Nurmio (eds.), *Palgrave handbook of Celtic languages and linguistics*. London: Palgrave Macmillan.

Willis, David. 2006. Against N-raising and NP-raising analyses of Welsh noun phrases. *Lingua* 116. 1807–1839. DOI: 10.1016/j.lingua.2004.09.004.

Part II

Phonology and phonetics

Chapter 6

Phonology of the Celtic languages

Michael Hammond
University of Arizona

This chapter offers a review of the phonological systems of the extant Celtic languages, including their consonant and vowel inventories as well as their salient phonological properties. The chapter focuses especially on how these phonologies differ from each other and on the similar developments that have occurred among them.

1 Introduction

In this chapter, I offer an overview of the synchronic phonology of the modern Celtic languages. My goal is to see what is common to these languages and to take seriously the proposition that there might be a more general Celtic phonological system that characterizes all of these languages together. The individual languages vary, of course, but I proceed on the hypothesis that the variations are in balance and that the differences between the languages aren't random but are tempered by other things.

The organization of this chapter is as follows. First, I lay out the general historical and geographic context for the modern Celtic languages. I then turn to a *very* brief overview of the phonological systems of each of the Celtic languages, focusing on points of similarity and comparison. I then try to synthesize these different systems and discuss the common threads that run through the phonologies of all the languages.

There are many many issues I simply cannot cover here given the breadth of the topic. Thus, my goal is to give enough of a sense of these individual systems to address the larger question, so that interested readers can pursue specific topics further on their own.

Michael Hammond. 2025. Phonology of the Celtic languages. In Andrew Carnie, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond & Luis Irizarry (eds.), *Foundational approaches to Celtic linguistics*, 211–245. Berlin: Language Science Press.
DOI: 10.5281/zenodo.15654867

2 Overview and context

Celtic is a branch of the Indo-European family. The modern representatives of this group include Breton, Cornish, Irish, Manx, Scottish Gaelic, and Welsh. Breton, Cornish, and Welsh comprise the Brythonic subgroup and Irish, Manx, and Scottish Gaelic the Goidelic subgroup (Fife 2010).

Celtic languages used to be spoken across much of Europe but now are spoken only in Ireland, Britain, and France (Eska 2010).[1] It's generally thought that Celtic peoples migrated from the European continent to the British islands and their languages then died out on the continent. The Breton language in France resulted from Celtic migration *back* to the continent.

Subsequent migrations or invasions by various Germanic peoples, Romans, and Norman French forced the various Celtic groups to the peripheries of Britain, Ireland, and France. In all cases, the languages are in significant and increasing competition with large national languages.

The speaker populations are all fairly small and difficult to quantify: Do we include all speakers or restrict ourselves to first language speakers? Do we only include speakers who use the language regularly? Do we expect speakers to be literate in their language? Table 1 gives *very* rough estimates of speaker populations for each language, based on a fairly generous interpretation of native speaker.[2]

Table 1: Estimated speakers

Language	Speakers
Breton	200,000
Cornish	0
Irish	40,000
Manx	0
Scottish Gaelic	50,000
Welsh	500,000

The numbers for Cornish and Manx require some context. Both of them are formerly dormant languages that are currently undergoing active revitalization

[1]There are, of course, expatriate communities in other places. The most significant ones are: Scottish Gaelic in Nova Scotia, Canada and Welsh in Patagonia, Argentina.

[2]More official, though perhaps less accurate, estimates are available at https://www.ethnologue.com.

efforts (Lowe 2018, Lewin 2022). The last first-language native speaker of Manx, Ned Maddrell, died in 1974. The last first-language native speaker of Cornish, Dolly Pentreath, died in 1777. Both languages have a number of second language speakers, probably numbering near 1000 in each case. In the case of Manx, there are reportedly new first-language native speakers. The first-language situation with Cornish is unknown at this time.

The Celtic languages exhibit a number of frequently cited typological features that are fairly rare (Fife 2010). These include the following:

1. consonant mutation
2. inflected prepositions
3. VSO word order

Other properties are occasionally cited as well, but these three are generally considered to be core Celtic properties. In order to get a sense of the larger grammatical setting within which each phonology is seated, each of these properties are investigated in turn.

The consonant mutations involve systematic changes to the initial consonant of a word in specific morphological or syntactic contexts. For example, in all the Celtic languages, the initial consonant of a feminine singular noun will change after the definite article.[3] This consonant change is overtly encoded in the orthographies of these languages, as seen in Table 2.

Table 2: Consonant mutation with feminine nouns after the definite article

Language	Noun	Gloss	With article
Breton	*bro*	'country'	*ar vro*
Cornish	*tus*	'people'	*an dus*
Irish	*bean*	'woman'	*an bhean*
Manx	*ben*	'woman'	*yn ven*
Scottish Gaelic	*caora*	'sheep'	*a' chaora*
Welsh	*cath*	'cat'	*y gath*

[3]The Goidelic languages distinguish grammatical case using the mutations; in those languages, the generalization applies only in the nominative.

The key fact is that words that begin with the same consonant mutate the same way in the same environment. Thus, for example, any feminine singular noun that begins with [b] in Breton mutates to [v] after the definite article. There are additional niceties in each language, which are discussed below.

All the modern Celtic languages exhibit inflected prepositions, where some prepositions are marked for the person and number of their objects. In Table 3 and Table 4 we see the relevant orthographic forms for the cognate preposition 'on'. These may or may not co-occur with the relevant overt pronoun. For example, in Scottish Gaelic, the overt pronoun does not co-occur with an inflected preposition; in Welsh, an overt pronoun does not occur in formal registers but typically does occur in the spoken language, e.g. *arna i* 'on me'.

Table 3: Inflected preposition 'on' in Breton, Cornish, and Irish

	Breton	Cornish	Irish
1SG	*warnon*	*warnav/m*	*orm*
2SG	*warnout*	*warnas*	*ort*
3SGM	*warnañ*	*warnadho*	*air*
3SGF	*warni*	*warnedhi*	*uirthi*
1PL	*warnomp*	*warnan*	*orainn*
2PL	*warnoc'h*	*warowgh*	*oraibh*
3PL	*warno*	*warnedhans*	*orthu*

Table 4: Inflected preposition 'on' in Manx, Scottish Gaelic, and Welsh

	Manx	SG	Welsh
1SG	*orrym*	*orm*	*arna*
2SG	*ort*	*ort*	*arnot*
3SGM	*er*	*air*	*arno*
3SGF	*urree*	*oirre*	*arni*
1PL	*orrin*	*oirnn*	*arnon*
2PL	*erriu*	*oirbh*	*arnoch*
3PL	*orroo*	*orra*	*arnyn*

The Celtic languages are also noted for VSO (verb–subject–object) word order. In all of them, this occurs when the verb is directly inflected (e.g., in the simple past tense). Following are examples in Welsh (1) and in Scottish Gaelic (2).

(1) Gwelodd y dyn y gath
saw the man the cat
'The man saw the cat.'

(2) Chunnaic an duine an cat
saw the man the cat
'The man saw the cat.'

As previously noted, Cornish and Manx are actively undergoing revitalization, and this surely has consequences for how researchers might characterize the

structures of these languages. However, it's fair to say that the other languages are also strongly affected by their contact with neighboring majority languages. To put it bluntly, natural conversations in these languages almost always involve code-switching or nonce borrowing with the neighboring dominant language.

For example, in speaking Welsh casually and talking about 'voting in the election', one is far more likely to hear borrowed or code-switched *fotio yn y lecsiwn* [vɔtjo ən ə lɛkʃʊn], rather than native *pleidleisio yn yr etholiad* [pʰlejdlejʃo ən ər ɛθɔljad].

This general situation has obvious consequences for the phonologies of these languages. The phonologies don't exist autonomously but in a, perhaps sometimes conflicting, relationship with other phonologies. For example, Bennett et al. (2018) discuss how the palatalization contrast is being lost in various Irish dialects, presumably as an effect of contact with English. Morris (2021) discusses a very interesting example where some Welsh varieties have acquired English [ɹ], but some English varieties have acquired Welsh [r, ɾ].

Keeping this in mind, I now turn to a very brief overview of the specific phonologies of each of the Celtic languages, focusing on inventory, mutation, and any salient phonological processes. The systems are treated in two groups. The Brythonic or P-Celtic languages can be opposed to the Goidelic or Q-Celtic languages. As noted above, the Brythonic languages are comprised of Breton, Cornish, and Welsh; the Goidelic languages include Irish, Manx, and Scottish Gaelic. The P- and Q- nomenclature reflects the fact that Indo-European *kʷ generally corresponds to Brythonic [p] and Goidelic [k] (written as either ⟨c⟩ or ⟨q⟩ in the orthography). Table 5 provides some examples of these correspondences in word-initial position.

Table 5: Reflexes of Indo-European *kʷ

	'children'	'five'	'what'	'four'
P-Celtic				
Breton	*(bugale)*	*pemp*	*pe*	*pevar*
Cornish	*(flehes)*	*pemp/pymp*	*pe/py*	*peswar*
Welsh	*plant*	*pump*	*pa*	*pedwar*
Q-Celtic				
Irish	*clann*	*cúig*	*cén*	*ceithre*
Manx	*cloan*	*queig*	*cre*	*kiare*
Scottish Gaelic	*clann*	*còig*	*(dè)*	*ceithir*

3 Goidelic

The phonologies of the Goidelic languages exhibit some commonalities. First, all the languages exhibit a palatalization contrast among the consonants, an innovation from Proto-Celtic. Second, the mutation systems are somewhat simpler than in the Brythonic languages and include at most two mutations, traditionally termed *lenition* and *eclipsis*.[4] On the other hand, the mutations affect more consonants and the mappings are somewhat more dramatic. Third, all the languages exhibit vowel reduction of stressless short vowels. Finally, the languages all include a marked epenthesis process involving clusters of a sonorant and an obstruent.[5]

3.1 Irish

Irish is spoken in Ireland, principally the Republic of Ireland, but to a lesser extent in Northern Ireland as well. The densest speaker populations are in the *Gaeltacht* or rural areas predominantly along the west coast of Ireland in the counties of Donegal, Galway, Kerry, and Mayo. In the south and east, Cork, Waterford, and Meath are official Gaeltacht areas as well. There are substantial numbers of second language speakers in other parts of the country, especially Dublin. There are also what we might term "new" speakers, speakers whose proficiency comes from immersion at school or in communities, but whose parents are not speakers. This is an important category for the other Celtic languages as well.

The language exhibits substantial dialect variation. Traditionally, three main dialect areas are distinguished.

- Munster (Cork, Kerry, Waterford)
- Connacht (Galway, Mayo)
- Ulster (Donegal)

There are interesting phonological differences between the dialects, but I will largely set these aside given the scope of the chapter.

There are excellent comprehensive phonological treatments of the language, including Ní Chiosáin (1991), Green (1997), Hickey (2014), and Ó Siadhail (1989). Generalizing across the dialects, the consonants are given in Table 6. Notice that

[4]This is the traditional analysis. Some analysts posit many more, e.g. Carnie (2008) for Irish.

[5]As I show later, the process appears to have occurred in Manx, but does not seem to be active in the synchronic language.

there is a contrast between plain (properly, velarized) and palatalized consonants for virtually all consonants. As well, in some dialects the consonants include velarized versions of some or all of [l, r, n] that, together with palatalization, makes for a four-way contrast among these segments.

Table 6: Irish consonants

			Palatalized				
	Labial	Coronal	Labial	Coronal	Dorsal	Dorsal	Glottal
Stop	p, b	t, d	p^j, b^j	t^j, d^j	k^j, g^j	k, g	
Nasal	m	n	m^j	n^j	$ŋ^j$	ŋ	
Fric.	f, v	s	f^j, v^j	s^j	x^j, j	x, ɣ	h
Approx.		l, r		l^j, r^j			

The Irish vowels appear in Table 7. Note that there is a contrast for length at all heights with some qualitative effects, though it's unclear whether this contrast extends to the low vowel(s). Ní Chiosáin (1991) proposes a much sparer input inventory for the short vowels.

Table 7: Irish vowels

	Front	Central	Back
High	i, iː		ʊ, uː
Mid	e, eː	ə	ɔ, oː
Low		a, aː	

Historically, the palatalization contrast for the consonants can be connected to the vowels, but the vowel system has undergone such radical changes that this is no longer the case for long vowels, and the contrast is phonemic synchronically. Short non-low vowels do show a relationship: they are front or back depending on the palatalization value of the following segment. Ní Chiosáin (1991: 141) cites the examples in Table 8.

There are a number of alternations between palatalized and plain consonants, but these are all morphologized. The only non-morphologized part of these alternations is that they are transitive. When a consonant becomes palatalized or non-palatalized, adjacent consonants to the left follow suit. In other words, sequences of adjacent consonants agree in palatalization.

Table 8: Backness of a non-low vowel determined by the following consonant in Irish

cur	[kur]	'put'	nonfin.
cuir	[kir^{j}]	'put'	imper.
muc	[muk]	'a pig'	nom. sg.
mhuic	[wik^{j}]	'a pig'	dat. sg.

Like all the other Celtic languages, Irish has initial consonant mutation. As previously noted, the initial consonant of a word undergoes systematic changes depending on its morphosyntactic environment. Setting aside the nature of the environment, traditionally there are two basic mutation patterns in Irish.[6] Following traditional terminology, these are referred to as *lenition* and *eclipsis*. Both are laid out in Table 9. Basically, lenition converts stops to fricatives; compare examples (3–4) and (5–6). In addition, voiceless fricatives either delete or debuccalize. Coronal consonants also debuccalize. Finally, [m] and [m^{j}] become fricatives as well. Eclipsis makes voiceless stops voiced and makes voiced stops into nasals. In addition, the fricatives [f] and [f^{j}] become voiced. These changes are overtly expressed in the orthography as well.[7]

As already noted, the mutations are triggered by morphosyntactic environments so they are curiously morphosyntactic and phonological at the same time. On the one hand, the morphosyntactic conditioning suggests that they should be viewed as something morphological. On the other hand, the fact that the consequences of mutation are phonological suggests that it should be viewed as a phonological process.

There is a virtual industry in debating which of these is correct for different mutations in the different Celtic languages from myriad theoretical perspectives. See, for example: Hamp (1951), Rogers (1972), Anderson (1976), Lieber (1992), Ball & Müller (1992), Kibre (1997), Pyatt (1997), Stewart (2004), Green (2006), Hannahs (2011a, 2013). I will not address this question here and instead focus on the phonological *consequences* of this process.

The most obvious consequence of these is that they alter the set of possible word-initial consonants and clusters (Grijzenhout 1995, Ní Chiosáin 1999). Thus, for example, [x] is not generally allowed word-initially in content words but is permitted word-initially in lenition contexts, as shown in the past tense forms (4) and (6).

[6]Though see Carnie (2008) for a more complex picture.

[7]Note that palatalization is expressed with vowel symbols in the spelling.

Table 9: Irish consonant mutation

Phonology			Orthography		
Base	Lenition	Eclipsis	Base	Lenition	Eclipsis
p	f	b	p	ph	bp
t	h	d	t	th	dt
t^j	h^j/x^j	d^j	t	th	dt
k	x	g	c	ch	gc
k^j	x^j	g^j	c	ch	gc
b	v	m	b	bh	mb
b^j	v^j	m^j	b	bh	mb
d	ɣ	n	d	dh	nd
d^j	j	n^j	d	dh	nd
g	ɣ	ŋ	g	gh	ng
g^j	j	$ŋ^j$	g	gh	ng
f	∅	v	f	fh	bhf
f^j	∅	v^j	f	fh	bhf
s	h	s	s	sh	s
s^j	$h^j/h/x^j$	s^j	s	sh	s
m	v	m	m	mh	m
m^j	v^j	m	m	mh	m

(3) caithim é
throw-1sg it
[kaim eː]
‘I throw it.’

(4) chaith mé é
threw I it
[xa(j) meː eː]
‘I threw it.’

(5) caillim é
lose-1sg it
[kal^{j}im eː]
‘I lose it.’

(6) chaill mé é
lost I it
[xalʲ meː eː]
'I lost it.'

Similarly, the mutations create word-initial clusters not otherwise allowed outside of mutation contexts. For example, while [sn] and [sʲlʲ] are generally legal onsets, [hn] and [hʲlʲ] are only legal in lenition contexts; they are not otherwise acceptable. Compare (7) and (9) with (8) and (10); again, all in the past tense:

(7) snámhaim
swim-1sg
[snaːwim]
'I swim.'

(8) shnámh mé
swam I
[hnɔv meː]
'I swam.'

(9) sleamhnaím
slip-1sg
[sʲlʲawniːm]
'I slip.'

(10) shleamhnaigh mé
slipped I
[hʲlʲawnə meː]
'I slipped.'

These sorts of effects on the inventory of (word-initial) segments and the inventory of word-initial consonant clusters run through all the Celtic languages.

Another very salient aspect of Irish phonology that has received a lot of attention in the literature is epenthesis into rising sonority clusters.[8] The phenomenon is curious in a number of respects. First, the clusters that this applies to are composed of a sonorant followed by a voiced consonant or a voiceless fricative; epenthesis is blocked when the second consonant is a voiceless stop or if the consonants are homorganic. The context following the cluster is irrelevant. In Table 10, I give examples where epenthesis occurs (from Ní Chiosáin 1999: 560).

[8]In the literature, this is sometimes referred to as *svarabhakti*.

In Table 11, I give examples where epenthesis does not occur because the second consonant is a voiceless stop (from Ní Chiosáin 1999: 561). And in Table 12, I give cases where epenthesis is blocked because the two consonants are homorganic.

Table 10: Examples of Irish epenthesis

lb	*bolb*	[boləb]	'caterpillar'
lw	*gealbhan*	[gʲaləwən]	'sparrow'
lx	*tulchach*	[tuləxəx]	'hilly'
rb	*borb*	[borəb]	'abrupt'
rm	*gorm*	[gorəm]	'blue'
rx	*dorcha*	[dorəxə]	'dark'
nb	*binb*	[bʲinʲibʲ]	'venom'
nm	*ainm*	[anʲimʲ]	'name'
nv	*leanbh*	[lʲanəv]	'child'

Table 11: Voiceless stop examples (no Irish epenthesis)

rp	*corp*	[korp]	'body'
rt	*gort*	[gort]	'field'
rk	*cearc*	[kʲark]	'hen'
lp	*spalp*	[spalp]	'burst forth'
lt	*alt*	[alt]	'joint'
lk	*olc*	[olk]	'evil'
nt	*cantal*	[kantəl]	'irritation'

Table 12: Homorganic consonant examples (no Irish epenthesis)

lr	*iolra*	[olrə]	'plural'
ld	*gallda*	[galda]	'foreign'
ls	*foilsigh*	[folʲsʲi]	'reveal'
mb	*siombalach*	[sʲombələx]	'symbolic'
nd	*gandal*	[gandal]	'gander'
ng	*teanga*	[tʲaŋgə]	'language'

Finally, the process also cannot occur when the preceding vowel is long. Ní Chiosáin (1991) cites these examples: *téarma* [tʲeːrmə] 'term' and *léargas* [lʲeːrgəs] 'insight'.

This is a devilishly tricky phenomenon to treat. First, the process exhibits no alternations. The first consonant must be a coda, but epenthesis doesn't care whether the second consonant is tautosyllabic or heterosyllabic. There are no processes that bear on the consonantal restrictions. Finally, the length of the preceding vowel doesn't alternate. The main evidence for the process is the surface absence of clusters where epenthesis must apply (but does not).

An additional argument can be made from the morphology (Ní Chiosáin 1999). When a final consonant is palatalized for morphological reasons, this palatalization also occurs on the consonant preceding the epenthetic vowel, as if the consonants were adjacent when palatalization occurs, e.g. in *colm* 'dove' [koləm], but *coilm* 'doves' [kelʲimʲ].

A second complication is that syllable structure doesn't matter, as shown by the irrelevance of what follows the cluster. Yet moraic status *does* matter as far as the preceding vowel (since it cannot be long). Finally, the odd restriction on voiceless stops presents an additional complication (Fullwood 2013). I return to the process of epenthesis further on because very similar facts obtain in Scottish Gaelic.

Stress in most dialects of Irish falls on the initial syllable, except for a few borrowings. However, stress in Munster Irish is quite striking and has received a fair amount of attention in the theoretical literature, e.g. Iosad (2013) and Bennett (2017).

In Munster, stress falls on the first syllable, e.g. in *clagarnach* [ˈklagərnəx] 'clattering'. If one of the first three syllables contains a long vowel or diphthong, then it gets stress, e.g. in *cailín* [kaˈlʲiːnʲ] 'girl', *amadántaíocht* [əməˈdɑːntiːxt] 'foolishness'.

There are two interesting additional complications. First, if the first two syllables are long, then stress falls on the second one, e.g. *díomhaoin* [dʲiːˈviːnʲ] 'idle'. Otherwise stress falls on the first long vowel in the first three syllables, e.g. *amparánaíocht* ['oumpərɑːniːxt] 'ungainliness'.

Second, if there are no long vowels or diphthongs in the first three syllables and the second vowel is [a] followed by [x], then it gets stressed, e.g. in *bacach* [bəˈkax] 'beggar', *bacacha* [bəˈkaxə] 'beggars', *casachtach* [kəˈsaxtəx] 'coughing'.

Finally, a comment on syllable structure. It's been observed that a single intervocalic consonant affiliates to the syllable to the left in Irish, rather than to the right. Ní Chiosáin et al. (2012) investigate this claim experimentally and find that, in fact, affiliation to the left is preferred when the preceding vowel is short, and

that the consonant typically affiliates in *both* directions. Strikingly, very similar facts have been cited for English (Hammond 1997; 1999). Notice that the stress facts just cited for Munster [ax] sequences lend support to such an account, as [ax] would constitute a syllable rhyme on this view.

3.2 Manx

The Manx language is another Goidelic language very closely related to Irish. As already noted above, Manx is spoken on the Isle of Man and went dormant upon the death of the last native speaker (at that time) in 1974. However, also as previously noted, Manx is currently the object of active revitalization efforts and there are now over a thousand speakers, including children.

As with Irish, I outline the segmental inventory, the mutations, and salient aspects of the phonology. Where appropriate, I compare these with the Irish system above. It's important to keep in mind, though, that Manx is in great flux due to the recent revitalization efforts. My description is based mostly on Broderick (1984).

The Manx consonants appear in Table 13. The system is quite similar to that of Irish, though note the absence of palatalized labials. Note also that the coronal stops are dental and occasionally affricated.

Table 13: Manx consonants

			Palatalized			
	Labial	Coronal	Coronal	Dorsal	Dorsal	Glottal
Stop	p, b	t͡θ, d͡ð	tʲ, dʲ	kʲ, gʲ	k, g	
Fricative	f, v	s	sʲ	xʲ	x	h
Nasal	m	n	nʲ	ŋʲ	ŋ	
Approximant	w	l, r	lʲ, j			

Manx has a simple five-vowel system with a length contrast and schwa. The vowels appear in Table 14.

There is a sporadic process of intervocalic weakening of stops, but it's not clear under what conditions this occurs and what the precise targets are. For example, a word like *scapail* 'escaping' can be pronounced [skəpʰeːlʲ], [skəβeːlʲ], or [skəveːlʲ]. Similarly, *brattag* 'flag' can be pronounced [bratag], [bradag], or [braðag]. Voiced stops exhibit a similar pattern. For example: *çhibbyrt* 'well' as [tʲubərt], [tʲuβərt], or [tʲuvərt]. The lenition mutation occurred historically

Table 14: Manx vowels

	Front	Central	Back
High	i, iː		u, uː
Mid	e, eː	ə	o, oː
Low		a, aː	

within words and it's reasonable to suppose this weakening process is a similar reassertion of that pattern.

Another salient process is prestopping where nasals and the liquid [l] can acquire a preceding voiced stop when the preceding vowel is stressed, as exemplified in Table 15. Interestingly, as we'll see later, a similar process occurs in Cornish.

Table 15: Manx prestopping

cam	'crooked'	[kam]	[kabm]
eeym	'butter'		[iːbm]
bane	'white'	[beːn]	[be(ː)dn]
keayn	'sea'		[kidn]
lhong	'ship'	[loŋ]	[lo/ugŋ]
shooyil	'walking'	[sʲuːl]	[sʲuːdl]
keeill	'church'	[kʲiːlʲ]	[kʲidlʲ]

Unlike the other Goidelic languages, Manx does not seem to exhibit epenthesis. There are many examples of forms that would undergo epenthesis in the other Goidelic languages that do not do so in Manx. See Table 16 for examples where no epenthesis appears in Manx but does appear in the corresponding Irish form. On the other hand, there are a number of cases where the cognate form in Manx has apparently undergone epenthesis (see Table 17). It would seem reasonable to suppose that Manx once had epenthesis, but it has since been lost leaving now lexicalized remnants.

Manx mutations are given in Table 18. It's important to keep in mind that eclipsis is virtually absent from the modern language.[9]

[9]Eclipsis occurs only after the article in the genitive plural and after the plural possessive pronoun *nyn*.

Table 16: No epenthesis in Manx

Manx			Irish	
jiarg	[dʲarg]	'red'	*dearg*	[dʲarəg]
pershagh	[pɛrʒəx]	'peach'		
argid	[ərgəd]	'money'	*airgead*	[ærʲɪgʲəd]
lorg	[lɔrg]	'pole'		
margher	[margɛ]	'market'		

Table 17: Lexicalized epenthesis in Manx

Manx			Irish	
gorrym	[gɔrəm]	'blue'	*gorm*	[gɔrəm]
dorraghey	[dɔrəxə]	'dark'	*dorcha*	[dɔrəxə]
ennym	[ɛnəm]	'name'	*ainm*	[ænʲɪm]
(thie) fillym	[fɪləm]	'cinema'		

3.3 Scottish Gaelic

Scottish Gaelic is the third Goidelic language.[10] The language formerly extended over much of Scotland, but is now fairly restricted to the extreme northern and northwestern peripheries and islands. There was substantial emigration during the highland clearances in the 18th and 19th centuries, but, except perhaps for Nova Scotia and Prince Edward Island in Canada, there are no current substantial speaker communities outside of Scotland.

The language is the object of active revitalization efforts and there are now Gaelic medium schools in Edinburgh and Glasgow, as well as in the north. Historically, there is a great deal of dialect variation, but this is greatly reduced as a function of language attrition and media in English. There are a number of useful discussions of the phonology and phonetics of Scottish Gaelic in the literature and my discussion draws on a number of these, for example: Bauer (2011), Bosch (1998), Bosch (1991), Bosch & de Jong (1997), Clements (1986), Green (1997), Ladefoged et al. (1998), Hammond et al. (2014), Hammond et al. (2017), Nance (2015), Smith (1999).

[10]The language is occasionally referred to as "Scots Gaelic" but this term is dispreferred by speakers of the language.

Table 18: Manx mutations

Base	Lenition	Eclipsis
p	f	b
t	h	d
t^j	h/x^j	d^j
k	x	g
k^j	x^j	g^j
b	v	m
bw	w	mw
m	v	
mw	w	
d	ɣ	n
d^j	j	n^j
g	ɣ	ŋ
g^j	j	$ŋ^j$
f	∅	v
fw	hw	w
s	h	
s^j	h/x^j	

The Scottish Gaelic consonants appear in Table 19. Note a palatalization contrast as in Irish and Manx. Note, too, the presence of a velarized series for sonorant coronals. These are not all present in all dialects and the contrast is occasionally realized instead in terms of length and/or diphthongization of the preceding vowel. The voicing contrast for stops is different from the other Goidelic languages. The contrast in initial position is between aspirated and unaspirated voiceless stops. Medially, the unaspirated series is typically voiced, and the aspirated series is typically *preaspirated* in non-initial position, e.g. in *cat* [k^ha^ht] 'cat'.[11]

The Scottish Gaelic vowels appear in Table 20. Note the presence of length and nasality contrasts, though the nasality contrast is not extensive and may be being lost among younger speakers. Note, too, that there are non-low back unrounded vowels in the system.[12]

[11]The word-initial version of this second series is what appears in Table 19.

[12]Chasaide (1995) reports that [ɤ] occurs in some varieties of Irish.

Table 19: Scottish Gaelic consonants

				Palatalized			
	Labial	Dental	Alveolar	Coronal	Dorsal	Dorsal	Glottal
Stop	pʰ, p	t̪ʰ, t̪		tʲʰ, tʲ	kʲʰ, kʲ	kʰ, k	
Nasal	m		n	ɲ		nˠ	
Fricative	f, v		s	ʃ	ç, ʝ	x, ɣ	h
Lateral			l	lʲ		l̪ˠ	
Rhotic			ɾ	ɾʲ		rˠ	
Approx				j			

Table 20: Scottish Gaelic vowels

	Front	Central	Back	
			Round	Unround
High	i, iː, ĩ, ĩː		u, uː, ũ, ũː	ɯ, ɯː, ɯ̃, ɯ̃ː
Mid High	e, eː		o, oː	ɤ, ɤː
Mid Low	ɛ, ɛː, ɛ̃, ɛ̃ː		ɔ, ɔː, ɔ̃, ɔ̃ː	
High		a, aː, ã, ãː		

The Scottish Gaelic lenition mutation is given in Table 21. The effects are essentially the same as in Irish and Manx. Note that in dialects with the velarization contrast, lenition may also affect these consonants such that sonorant coronals lose velarization and palatalization in the relevant contexts. When this occurs, it is not represented in the orthography.

Eclipsis mutation is absent in many dialects and not reflected in the orthography. If it does occur, it has the effects in Table 22. The expected pattern is shown in the Eclipsis column. However, Ó Maolalaigh (1995) notes additional patterns, which appear in different dialects; these are shown in the right three columns.[13]

Another interesting difference between Scottish Gaelic and the other Goidelic languages is that the trigger for eclipsis always ends in a nasal consonant. This

[13] Ó Maolalaigh describes the patterns in terms of a voicing contrast; I've translated it into the aspiration contrast used here. It's a little unclear what Ó Maolalaigh intends by the voiced aspirated symbols. The key point about the B group is that a voicing distinction of some sort is maintained under eclipsis.

Table 21: Scottish Gaelic lenition mutation

Orthographic		Plain		Palatalized	
Unlenited	Lenited	Unlenited	Lenited	Unlenited	Lenited
p	ph	p^h	f	—	—
b	bh	p	v	—	—
t	th	t^h	h	t^{jh}	ç
d	dh	t	ɣ	t^j	ʝ
c	ch	k^h	x	k^{jh}	ç
g	gh	k	ɣ	k^j	ʝ
f	fh	f	∅	—	—
s	sh	s	h	ʃ	ç
m	mh	m	v	—	—

Table 22: Scottish Gaelic mutation

Consonant	Eclipsis	A	B	C
p^h	p	p	b^h	m^h
t^h	t	t	d^h	n^h
k^h	k	k	g^h	$ŋ^h$
p	m	p	p	m
t	n	t	t	n
k	ŋ	k	k	ŋ
f	v	f	f	

may be evidence that eclipsis is in the process of becoming phonological again in the language.

Finally, Scottish Gaelic exhibits essentially the same epenthesis process as Irish. See Table 23 for a few examples. Note that the inserted vowel (bolded) is a copy of the preceding vowel in Scottish Gaelic. In addition, there are interesting prosodic properties associated with the inserted vowel in Scottish Gaelic (Bosch & de Jong 1997, Bosch 1998). For example, Bosch & de Jong (1997) report that the epenthetic vowel in Barra Scottish Gaelic is acoustically stressed.

The process is blocked in the same contexts as in Irish as shown in Table 24: when the first consonant is not a sonorant, when the preceding vowel is long or a

Table 23: Epenthesis (bolded) in Scottish Gaelic

seanmhair	[ʃɛnɛvɛr]	‘grandmother’
arm	[ara**ə**m]	‘army’
dorcha	[dɔrɔxə]	‘dark’
tilg	[t͡ʃɪlɪk]	‘to toss away’
urchair	[ʊr**ʊ**xʊr]	‘shot’

diphthong, when the consonants are homorganic, or when the second consonant is a voiceless stop.

Table 24: Scottish Gaelic examples where epenthesis is blocked (marked with underline)

Nonsonorant C1	*smachd*	[smax_k]	‘stifle’
	sgriobhte	[skriv_t͡ʃə]	‘written’
Long vowel	*dualchas*	[duəl_xəs]	‘tradition’
	iarmailt	[iːrmɛl_t͡ʃ]	‘firmament’
Homorganic CC	*Òlaind*	[oːlan_d͡ʒ]	‘Holland’
	mandrag	[man_drak]	‘mandrake’
C2 Voiceless Stop	*olc*	[ɔl̥_k]	‘evil’
	corp	[kɔr̥_p]	‘body’

Note how the restriction against the second consonant being a voiceless stop makes a bit more sense in Scottish Gaelic than in Irish. In Scottish Gaelic, medial voiceless stops are preaspirated. When there is a preceding sonorant, that sonorant is thus voiceless. We may then subsume the voiceless stop and homorganicity restriction under the same rubric of shared features blocking an intervening inserted element.

4 Brythonic

The Brythonic languages (Breton, Cornish, and Welsh) differ, broadly speaking, from the Goidelic languages in several respects.

1. None of the Brythonic languages make a distinction between broad and slender contrasts (a palatalization contrast). Concomitantly perhaps, none

of the Brythonic languages distinguish overt grammatical case, since this distinction is one of the markers.

2. The mutation systems in the Brythonic languages are richer and typically more direct in the phonological relationships they exhibit.

3. Except perhaps for late Cornish, none of the Brythonic languages have the kind of extensive vowel reduction evident in the Goidelic languages. Stress in Goidelic is either on the first syllable or may, in some cases and dialects, shift to a heavy syllable to the right. In the Brythonic languages, stress is typically penultimate.

4. The Brythonic languages do not exhibit the vowel epenthesis pattern we saw in Irish and Scottish Gaelic (though a different epenthesis pattern is evident in Welsh).

4.1 Breton

Breton is spoken in Brittany, France. It is unique among the Celtic languages in that it is in contact with French rather than English. However, the situation could be viewed as somewhat more complex. First, Breton speakers are thought to have migrated from Cornwall (and elsewhere) in the Middle Ages and thus, to that point, the language was under the influence of English (and its forebearers) and Latin. Second, another factor that sets it apart is that in its present setting in France, it is in some contact with the Gallo language, a Romance language closely related to French and spoken in upper Brittany.[14]

Breton exhibits considerable dialect variation and my characterization here attempts to generalize beyond these distinctions. Fuller descriptions of the general system can be found in Hemon (1995) and Press (2011).

The Breton consonant inventory appears in Table 25. Note the absence of a plain versus palatalized contrast and the presence of labio-velar stops.

It's worth noting that the rhotics [r, ɹ, ʁ] are extremely variable with some speakers producing taps and other assimilating fully to a French uvular. An unresolved issue is whether some obstruents contrast in fortis-lenis and/or length in at least some dialects.[15]

[14]Interestingly, Cornish shows evidence of French influence as well, at the very least because of the Norman conquest, but also because of continuing contact with Breton (Ferdinand 2013). Thus, one would not want to say that French influence on Breton began with its transposition to France.

[15]See Press (2011) for discussion.

Table 25: Breton consonants

	Labial	Coronal	Pal-Alv	Dorsal	Lab-Dors	Uvular/Glottal
Stop	p, b	t, d		k, g	k^w, g^w	
Fricative	f, v	s, z	ʃ, ʒ	x, ɣ		h
Nasal	m	n	ɲ			
Approx		l, r, ɹ	ʎ, j, ɥ		w	ʁ

A very salient consonant process in the language is final devoicing and voicing assimilation. Obstruents are voiceless in absolute phrase-final position but assimilate in voicing to a following segment. Compare (11) with (12–15) from Hemon (1995: 79).

(11) sellit ouzh ar beleg [sɛlid uz aʁ bɛːlɛk]
look at the priest
'Look at the priest.'

(12) setu amañ ar beleg kozh [sɛty amã aʁ bɛːlɛk koːs]
voila here the priest old
'Here is the old priest.'

(13) beleg ar barrez [bɛːlɛg aʁ baʁes]
priest the parish
'the priest of the parish'

(14) ur beleg mat [œʁ bɛːlɛg maːt]
a priest good
'a good priest'

(15) ur beleg mat eo [œʁ bɛːlɛg maːd eɔ]
a priest good is
'He is a good priest.'

Strikingly, if two identical obstruents come together, they surface as voiceless (as shown in 16 and 17; again, from Hemon 1995). This happens whether they initially agree in voicing or not.

(16) bloas zo [blwa(s) so]
year ago
'a year ago'

(17) dek gwele [de:k kwe:le]
10 beds
'10 beds'

However, the outcome does vary by dialect when both consonants are initially voiceless (Le Dû 1986). Interestingly, this same restriction is also part of Welsh classical meter, though it is *not* a part of the synchronic language (Morris-Jones 1980, ap Dafydd 1994, Llwyd 2007).

The Breton vowel inventory appears in Table 26. Note the presence of front round vowels and a contrast between open and close among mid vowels. Note too the robust nasalization contrast.

Table 26: Breton vowels

	Front		Back	
	Unround	Round	Unround	Round
High	i, ĩ	y, ỹ		u, ũ
Close-mid	e, ẽ	ø, ø̃		o, õ
Open-mid	ɛ, ɛ̃	œ, œ̃		ɔ, ɔ̃
Low	a, ã		ɑ, ɑ̃	

The Breton mutations appear in Table 27. The pattern is similar across the Brythonic languages: more mutations and a more transparent phonological relationship between each sound and what it mutates to.

There are a number of interesting restrictions on the Breton mutation system. Some are reminiscent of those seen in the Goidelic systems, but some are unique to Breton. For example, *va* [ɛm] 'my' generally triggers the aspirate mutation; compare (18–19) and (20–21). However, when it contracts to *m* with a previous vowel-final particle, it only triggers that mutation on non-labials; compare (22) with (23).

(18) kambr
room
[kambʁ]
'room'

Table 27: Breton mutations

	Mutation type			
C	Soft	Aspirate	Hard	Mixed
p	b	f		
t	d	z		
k	g	x		
b	v		p	v
d	z		t	t
g	ɣ		k	ɣ
gw	w		kw	w
m	v			v

(19) em c'hambr
to-my room-asp
[ɛm xambʁ]
'to my room'

(20) ti
house
[ti]
'house'

(21) em zi
to-my house-asp
[ɛm zi]
'to my house'

(22) paner
basket
[panɛr]
'basket'

(23) 'm paner
my basket
[m panɛr]
'my basket'

This is similar to the blocking of lenition that occurs with coronals in Irish and Scottish Gaelic (Goidelic languages). Table 28 gives examples of this lenition

process following a definite article that precedes a feminine singular noun. If the noun begins with a coronal stop, it does not mutate.[16]

Table 28: Lenition mutation in Irish and Scottish Gaelic. Bolded examples are cases where lenition is blocked before coronals.

Irish	*an bhean*	[a vaːn]	'the woman'
	an ghal	[a ɣal]	'the vapor'
	an deoch	**[a dʲɔx]**	'the drink'
Scottish Gaelic	*a' bhean*	[ə vɛn]	'the woman'
	a' chaileag	[ə χalak]	'the girl'
	an dànachd	**[ən daːnaxk]**	'the poetry'

4.2 Cornish

As already noted, Cornish became dormant upon the death of the last (at that time) native speaker in 1777. Nonetheless, it has been the subject of almost constant revitalization attempts since then. Interestingly, several different versions of modern Cornish exist, based either on middle or late Cornish. I view these simply as dialects, or better yet *chronolects*, of the same language. My main sources on the language are Jenner (2010) and Brown (2001).

The Cornish consonant inventory appears in Table 29. Note the absence of a plain versus palatalized contrast just as in Breton.

Table 29: Cornish consonants

	Labial	Dental	Alveolar	Pal-Alv	Dorsal	Glottal
Stop/Affricate	p,b	t,d		t͡ʃ,d͡ʒ	k,g	
Fricative	f,v	θ,ð	s,z	ʃ,ʒ	x	
Nasal	m	n			ŋ	
Approx	w,ʍ		l,ɾ/ɹ	j		h

Late Cornish undergoes an interesting pre-stopping process that is reminiscent of that seen in Manx, except that it is reflected in the orthography. Pre-stopping in Cornish occurs after a stressed vowel and before a nasal consonant,

[16]It is unclear whether this occurs in Manx (Teare & Wheeler 2017).

e.g. *kabm* [kabm] ‘step’ and *pedn* [pɛdn] ‘head’. If the nasal occurs in a stressless syllable, pre-stopping does not occur, e.g. *perghen* [ˈpɛrhɛn] ‘owner’ and *pedrevan* [pəˈdrɛvən] ‘lizard, newt’.

Since stress shifts under suffixation, this creates alternations in the plural, e.g. in *kabmas* [ˈkabmas] ‘bay, bend’ vs. *kamajow* [kaˈmad͡ʒow] ‘bays, bends’ and *pednyn* [ˈpɛdnɪn] ‘tadpole’ vs. *penydnow* [pɛˈnɪdnow] ‘tadpoles’.

There is evidence that the language used to undergo voicing assimilation and final devoicing as in Breton, but it’s unclear whether this occurs in the synchronic language.

The language has a number of sonority-reversed consonant clusters. These undergo epenthesis in final position, which is transparent in the Cornish orthography, but alternations occur in suffixed forms (for example, in the plural; see Table 30).[17]

Table 30: Epenthesis alternations in singular vs. plural forms in Cornish

vr	*lever*	[lɛvɐr]	‘book’
	levrow	[lɛvrow]	‘books’
tr	*ownter*	[ɔʊntɐr]	‘uncle’
	owntres	[ɔʊntrəs]	‘uncles’
dhl	*kenedhel*	[kɛnəðɐl]	‘nation’
	kenedhlow	[kənɛðlow]	‘nations’
ml	*amal*	[æmɐl]	‘border, side’
	emlow	[ɛmlow]	‘borders, sides’

Orthographically, the epenthetic vowel is a copy of the preceding vowel, but this does not seem to be the case in the spoken form. That this is truly epenthesis, rather than deletion, is evidenced by forms that don’t alternate; e.g. *lether* [lɛθɐr] ‘letter’ vs. *letherow* [ləθɛrow] ‘letters’.

The Cornish vowel inventory appears in Table 31. Note the length contrast and the presence of front rounded vowels. The latter only occur in middle Cornish and the corresponding chronolect.

Finally, the Cornish mutations appear in Table 32. Note that they are quite similar to the mutations in Breton.

[17] Bennett (2012) argues for a similar process in Irish. See Smith (1999) for Scottish Gaelic.

Table 31: Cornish vowels

	Front		Central	Back
	Unround	Round		
High	iː, ɪ	(yː, ʏ)		uː, ʊ
Mid	eː, ɛ	(øː, œ)	ə	oː, ɔ/ɤ
Low	æː, æ/a			ɑː, ɑ

Table 32: Cornish mutations

	Mutation Type			
C	Soft	Aspirate	Hard	Mixed
p	b	f		
t	d	θ		
k	g	h		
b	v		p	f
d	ð		t	t
g	∅,w		k	h
gw	w		kw	hw
m	v			f
t͡ʃ	d͡ʒ			

4.3 Welsh

Unsurprisingly, Welsh is spoken primarily in Wales, but there are also Welsh-speaking communities abroad, especially in England. Notably, there is a Welsh-speaking community in Argentina, where the Welsh established a colony to escape the effects of contact with the English (Bell 2018). Like the other Celtic languages, there are a number of dialects (Jones 1998, Thomas 2000, Awbery 1986), but there is a significant divide between dialects in the north and south, which have interesting phonological differences.

There is a fairly significant literature on Welsh grammar and phonology. This ranges from traditional grammars like Morris-Jones (1913) and King (2003) to more modern treatments like Hannahs (2013). There are more specific discussions of the mutation system that include Awbery (1973) and Hannahs (2011a) as well as a wonderful set of volumes that includes papers on a number of aspects

of the phonology, phonetics, and mutation system: Ball & Williams (2001), Ball & Müller (1992), and Ball & Jones (1984). Additional papers of note on other aspects of the sound system of Welsh include Hannahs (1996, 2009, 2011b), Hammond (2019), Iosad (2017b), Jones (1984), and Awbery (1984), among others.

The surface consonant inventory for Welsh is given in Table 33. Again, note the absence of a palatalization contrast. Unlike Breton and Cornish, there are no labio-velars, but it's not clear that this is a substantive difference. Cognate elements exist in Welsh, but are typically analyzed as a sequence of a velar followed by [w], e.g. *cwyno* [kʰwɨno] 'complain', *gwin* [gwin] 'wine', and *chwarel* [χwarɛl] 'quarry'.[18] Some researchers recognize a series of voiceless nasals as well, but I do not represent those here (Hammond 2019). Finally, note that the back unrounded glide [ɯ̯] only occurs in northern dialects.

Table 33: Welsh consonants

	Labial	Dental	Alveolar	Pal-Alv	Dorsal	Uvular/Glottal
Stop/Affric	pʰ, b	tʰ, d		t͡ʃʰ, d͡ʒ	kʰ, g	
Fricative	f, v	θ, ð	s, ɬ	ʃ		χ
Nasal	m	n			ŋ	
Approx	w		l, r, r̥	j	ɯ̯	h

I note two interesting phonological processes. First, as in Cornish, there are final sonority-reversed clusters that are resolved by epenthesis. Unlike in Cornish, however, there are two additional restrictions. First, per Hannahs (2009, 2011b), epenthesis is at least partially conditioned by stress. Sonority-reversed clusters in words of one syllable undergo epenthesis, e.g. *llyfr* ['ɬɨvɨr] 'book' and *pobl* ['pʰɔbɔl] 'people'. However, in longer words where the cluster appears in a stressless syllable, deletion of the final consonant is a more typical strategy, e.g. *ffenestr* ['fɛnɛst] 'window' and *posibl* ['pʰɔsib] 'possible'. Furthermore, the final clusters [dl] and [vn] do not seem to undergo either epenthesis or deletion, e.g. *cenedl* ['kʰɛnɛdl] 'nation' and *ofn* [ɔvn] 'fear', as noted in Hannahs (2013) and Iosad (2017b).

It's interesting to compare this pattern with Cornish. There, the cognate for Welsh *cenedl* does undergo epenthesis: *cenedhel*. However, the cognate form for Welsh *ofn* is Cornish *own* [ɔʊn], where the consonant (pronounced in Welsh as

[18] Evidence against this comes from distributional observations. For example, in unmutated onset position, [χ] is almost always followed by [w]. The only exception is *chi* [χi] 'you', historically *chwi* [χwi]. Similarly, while velars can be followed by [w], other segment classes cannot be.

[v]) is lost. The consonant [v] is also lost in Welsh, but at the ends of words, as in *haf* [ha(v)] 'summer', *cyntaf* [kʰəntʰa(v)] 'first', etc. Thus, the absence of epenthesis in Welsh forms like *ofn* may be due to the precarious status of [v].

Another prominent process in Welsh involves the deletion of a medial [h] when it precedes a stressless vowel (Hammond 2019). This process also affects [r̥], with its voiced counterpart [r] occurring where [h] would be deleted. Since, as in Cornish, stress moves under suffixation, this again gives rise to alternations, e.g. *dihareb* [diˈharɛb] 'proverb' vs. *diarhebion* [diaˈr̥ɛbjɔn] 'proverbs', *eang* [ˈeaŋ] 'wide' vs. *ehangach* [eˈhaŋaχ] 'wider'. The *dihareb* form is particularly nice as it shows both the h/∅ and the r/r̥ alternations.

Historically, medial nasal–voiceless stop sequences were replaced with nasal–[h] sequences; these exhibit alternations as well: *cenedl* [ˈkʰɛnɛdl] 'nation' vs. *cenhedloedd* [kʰɛnˈhɛdlɔ(ɯ)ð] 'nations', *angen* [ˈaŋɛn] 'need' vs. *anghenion* [aŋˈhɛnjɔn] 'needs', *tymor* [ˈtʰəmɔr] 'season' vs. *tymhorau* [tʰəmˈhɔra(ɯ)] 'seasons'.

The surface inventory of Welsh vowels is given in Table 34. Note that the high central unrounded vowels only occur in the north.

Table 34: Welsh vowels

	Front	Central	Back
High	iː, ɪ	ɨː, ɨ	uː, ʊ
Mid	eː, ɛ	ə	oː, ɔ
Low		aː, a	

Vowel length is only contrastive in monosyllables and then only before certain consonants. Historically, this only included [l, r, n], but this has been extended in recent borrowings where long vowels may now occur where previously only short ones did, e.g. *strôc* [stroːk] 'stroke, shock', *côt* [kʰoːt] 'coat', *grŵp* [gruːp] 'group', *clòs* [kʰloːs] 'close', etc.[19]

The Welsh mutations appear in Table 35. Note that the system is not as complex as in Breton and Cornish, but still more complex than in any of the Goidelic languages. As in the other Brythonic languages, the relationships are fairly transparent phonologically. Interestingly, similar to eclipsis in the Scottish Gaelic dialects cited above, the nasal mutation can be treated phonologically, triggered only by elements that end in an overt nasal (Hammond 2019).

[19]The distribution of vowel length in Welsh is discussed in Awbery (1984), Hannahs (2013), Iosad (2017a), and Bell (2018).

Table 35: Welsh mutations

	Mutation Type		
C	Soft	Nasal	Aspirate
p	b	mh	f
t	d	nh	θ
k	g	ŋh	χ
b	v	m	
d	ð	n	
g	∅	ŋ	
ɬ	l		
r̥	r		
m	v		

Finally, the stress system of Welsh deserves mention. Stress falls generally on the penult. There are a fair number of words with exceptional final stress and a very few items with antepenultimate stress. What's striking about the Welsh system is that English speakers routinely intuit stress as falling on the final syllable. There are two reasons for this. First, Welsh does not have stress-related vowel reduction so final stressless syllables are routinely not reduced. Second, as noted by Williams (1989) and Dogil & Williams (1999), the intonational system can, in some cases, give conflicting cues so that the final syllable surfaces with a fairly high pitch even when it isn't stressed.

5 General discussion

Table 36 provides a broad summary of interesting differences that emerge across the Celtic languages with respect to the phenomena discussed in this chapter.

The first row of Table 36 gives the number of mutations, showing that generally there are more of them in Brythonic than in Goidelic languages. The second row gives a crude characterization of the transparency of the mutations, where the mutations of the Goidelic languages can be seen as generally less transparent than those of the Brythonic languages. And, though it's not shown in Table 36, the Goidelic mutations also affect more consonants than those in the Brythonic languages. The next two rows show that the Goidelic languages generally exhibit vowel reduction and initial stress, whereas the Brythonic languages exhibit

Table 36: Summary of differences

	Goidelic			Brythonic		
	Irish	Manx	SG	Breton	Cornish	Welsh
Number of Mutations	2	1–2	2	4	4	3
Transparent Mutation	no	no	no	yes	yes	yes
V Reduction	yes	yes	yes	no	no	no
Stress Placement	initial*	initial	initial	penult	penult	penult
Grammatical Case	yes	?	yes	no	no	no
Palatalization Contrast	yes	yes	yes	no	no	no

penult stress with no or less vowel reduction. Finally, the last two rows show that overt grammatical case occurs in the Goidelic languages along with a palatalization contrast, while neither occur in the Brythonic languages.

Additionally, there are other commonalities described in this chapter that do not follow the Goidelic/Brythonic divide. For example, in all the languages, mutation is triggered by morphosyntactic environments. However, I also noted several cases of 'phonological mutation', namely the eclipsis mutation in Scottish Gaelic and the nasal mutation in Welsh. Finally, there is pre-stopping in virtually the same environment in two of the languages, Manx and Cornish.

Taking a closer look at the differences summarized in Table 36, one can observe that the phonology of the mutations is somewhat simpler in the Brythonic languages (as measured by transparency) than in the Goidelic languages – both in terms of the number of consonants affected and in terms of the mappings that occur. It would seem reasonable to suppose that this is balanced against the number of mutations in the Brythonic languages.

Another observation to ponder is that all the Goidelic languages exhibit a palatalization contrast whereas the Brythonic languages do not. One could ask if the absence of overt case in the Brythonic languages is connected to this, because case in the Goidelic languages is at least partially marked by palatalization.

Another potential point of covariance is vowel reduction and palatalization. The Goidelic languages exhibit both a palatalization contrast and vowel reduction whereas the Brythonic languages exhibit neither. Perhaps the way to look at these is a shifting of contrast from vowels to consonants in the Goidelic languages.

Of course, these observations are just hypotheses at this stage. To truly test them, these properties could be examined in the context of historical change and

language acquisition. For example, if these variables are connected as speculated above, I would expect them to arise diachronically in close succession. For example, is there evidence that vowel reduction arises when the stress systems separate? Similarly, I would expect to see these connections recapitulated in the timing of acquisition. For example, does the acquisition of the case system in the Goidelic languages correlate with the acquisition of the palatalization contrast?

In sum, these languages exhibit a range of interesting phonological properties. It is reasonable to see the variation across the systems as different ways of balancing phonological and morphological requirements and options. I've laid out several such possible connections in this chapter and propose that they can be tested with further study in the historical and language acquisition domains.

Acknowledgments

Thanks to Elise Bell, Ryan Bennett, Andrew Carnie, S.J. Hannahs, Diane Ohala, and two anonymous reviewers for useful discussion. All errors are my own.

References

Anderson, Stephen R. 1976. On the description of consonant gradation in Fula. *African Linguistics* 7. 93–136.

ap Dafydd, Myrddin. 1994. *Clywed cynghanedd: Cwrs cerdd dafod*. Llanrwst: Gwasg Carreg Gwalch.

Awbery, Gwenllian M. 1973. Initial mutation in a generative grammar of Welsh. *Phonetics Department Report* 4. University of Leeds, 28–46.

Awbery, Gwenllian M. 1984. Phonotactic constraints in Welsh. In Martin J. Ball & Glyn E. Jones (eds.), *Welsh phonology*, 6–104. Cardiff: University of Wales Press.

Awbery, Gwenllian M. 1986. *Pembrokeshire Welsh: A phonological study*. Wales: National Museum of Wales.

Ball, Martin J. & Glyn E. Jones (eds.). 1984. *Welsh phonology*. Cardiff: University of Wales Press.

Ball, Martin J. & Nicole Müller. 1992. *Mutation in Welsh*. London: Routledge.

Ball, Martin J. & Briony Williams. 2001. *Welsh phonetics*. Lewiston: Edwin Mellen Press.

Bauer, Michael. 2011. *Blas na gàidhlig*. Glasgow: Akerbeltz.

Bell, Elise. 2018. *Perception and production of Welsh vowels by Welsh-Spanish bilinguals*. University of Arizona. (Doctoral dissertation).

Bennett, Ryan. 2012. *Foot-conditioned phonotactics and prosodic constituency*. UC Santa Cruz. (Doctoral dissertation).

Bennett, Ryan. 2017. Output optimization in the Irish plural system. *Journal of Linguistics* 53. 229–277. DOI: 10.1017/s002222671500033x.

Bennett, Ryan, Máire Ní Chiosáin, Jaye Padgett & Grant McGuire. 2018. An ultrasound study of Connemara Irish palatalization and velarization. *Journal of the International Phonetic Association* 48. 261–304. DOI: 10 . 1017 / s0025100317000494.

Bosch, Anna. 1991. *Phonotactics at the level of the phonological word*. University of Chicago. (Doctoral dissertation).

Bosch, Anna. 1998. The syllable in Scottish Gaelic dialect studies. *Scottish Gaelic Studies* 18. 1–22.

Bosch, Anna & Kenneth de Jong. 1997. The prosody of Barra Gaelic epenthetic vowels. *Studies in the Linguistic Sciences* 27. 1–15.

Broderick, George. 1984. *A handbook of late spoken Manx*. Tübingen: Niemeyer.

Brown, Wella. 2001. *A grammar of modern Cornish*. Callington, Cornwall: The Cornish Language Board.

Carnie, Andrew. 2008. *Irish nouns*. Oxford: Oxford University Press.

Chasaide, Ailbhe Ní. 1995. Irish. *Journal of the International Phonetic Association* 25. 34–39.

Clements, George Nickerson. 1986. Syllabification and epenthesis in the Barra dialect of Gaelic. In Koen Bogers, Harry van der Hulst & Maarten Mous (eds.), *The phonological representation of suprasegmentals (studies dedicated to J. M. stewart)*, 317–336. Dordrecht: Foris Publications.

Dogil, Grzegorz & Briony Williams. 1999. The phonetic manifestation of word stress. In Harry van der Hulst (ed.), *Word prosodic systems in the languages of Europe*, 273–334. Berlin, New York: De Gruyter Mouton.

Eska, Joseph F. 2010. The emergence of the Celtic languages. In Martin J. Ball & Nicole Müller (eds.), *The Celtic languages*, 22–27. London: Routledge.

Ferdinand, Siarl. 2013. A brief history of the Cornish language, its revival and its current status. *e-Keltoi* 2. 199–227.

Fife, James. 2010. Typological aspects of the Celtic languages. In Martin J. Ball & Nicole Müller (eds.), *The Celtic languages*, 1–21. London: Routledge.

Fullwood, Michelle Alison. 2013. The perceptual dimensions of sonority-driven epenthesis. *Proceedings of the Annual Meetings on Phonology*.

Green, Antony Dubach. 1997. *The prosodic structure of Irish, Scots Gaelic, and Manx*. Cornell University. (Doctoral dissertation).

Green, Antony Dubach. 2006. The independence of phonology and morphology: The Celtic mutations. *Lingua* 116. 1946–1985. DOI: 10.1016/j.lingua.2004.09.002.

Grijzenhout, Janet. 1995. *Irish consonant mutation and phonological theory*. OTS Dissertation Series. University of Utrecht. (Doctoral dissertation).

Hammond, Michael. 1997. Vowel quantity and syllabification in English. *Language* 73. 1–17. DOI: 10.2307/416591.

Hammond, Michael. 1999. *The phonology of English*. Oxford: Oxford University Press.

Hammond, Michael. 2019. Voiceless nasals in Welsh. *Journal of Celtic Linguistics* 20. 31–60. DOI: 10.16922/jcl.20.3.

Hammond, Michael, Yan Chen, Elise Bell, Andrew Carnie, Diana Archangeli, Adam Ussishkin & Muriel Fisher. 2017. Phonological restrictions on lenition in Scottish Gaelic. *Language* 93. 446–472. DOI: 10.1353/lan.2017.0020.

Hammond, Michael, Natasha Warner, Andréa Davis, Andrew Carnie, Diana Archangeli & Muriel Fisher. 2014. Vowel insertion in Scottish Gaelic. *Phonology* 31. 123–153. DOI: 10.1017/s0952675714000050.

Hamp, Eric P. 1951. Morphophonemes of the Keltic mutations. *Language* 27. 230–247.

Hannahs, S.J. 1996. Phonological structure and soft mutation in Welsh. In Ursula Kleinhenz (ed.), *Interfaces in phonology*, 46–59. Berlin: Akademie Verlag.

Hannahs, S.J. 2009. Welsh svarabhakti: Sonority sequencing and foot structure. *Journal of Celtic Linguistics* 13. 21–44.

Hannahs, S.J. 2011a. Celtic mutations. In Marc van Oostendorp, Colin Ewen, Keren Rice & Elizabeth Hume (eds.), *The Blackwell companion to phonology*, vol. 5, 2807–2830. Malden, MA: Wiley-Blackwell.

Hannahs, S.J. 2011b. Unity in diversity in Welsh: The avoidance of sonority sequencing violations. In Andrew Carnie (ed.), *Formal approaches to Celtic linguistics*, 259–374. Newcastle upon Tyne: Cambridge Scholars.

Hannahs, S.J. 2013. *The phonology of Welsh*. Oxford: Oxford University Press.

Hemon, Roparz. 1995. *Breton grammar*. translated and revised by Michael Everson. Cathair na Mart: Evertype.

Hickey, Raymond. 2014. *The sound structure of Modern Irish*. Berlin: de Gruyter Mouton.

Iosad, Pavel. 2012. *Representation and variation in substance-free phonology: A case study in Celtic*. University of Tromsø. (Doctoral dissertation).

Iosad, Pavel. 2013. Head-dependent asymmetries in Munster Irish prosody. *Nordlyd* 40. 66–107. DOI: 10.7557/12.2502.

Iosad, Pavel. 2017a. The phonologisation of redundancy: Length and quality in Welsh vowels. *Phonology* 34. 121–162. DOI: 10.1017/s0952675717000057.

Iosad, Pavel. 2017b. Welsh svarabhakti as stem allomorphy. *Transactions of the Philological Society* 115. 141–175. DOI: 10.1111/1467-968x.12085.

Jenner, Henry. 2010. *Handbook of the Cornish language*. Revised by Michael Everson. London: Evertype.
Jones, Glyn E. 1984. The distinctive vowels and consonants of Welsh. In Martin J. Ball & Glyn E. Jones (eds.), *Welsh phonology*, 40–64. Cardiff: University of Wales Press.
Jones, Mari C. 1998. *Language obsolescence and revitalization: Linguistic change in two sociolinguistically contrasting Welsh communities*. Oxford: Clarendon Press.
Kibre, Nicholas J. 1997. *A model of mutation in Welsh*. Bloomington, IN: Indiana University Linguistics Club Publications.
King, Gareth. 2003. *Modern Welsh: A comprehensive grammar*. London: Routledge.
Krämer, Martin. 2000. Voicing alternations and underlying representations: The case of Breton. *Lingua* 110. 639–663. DOI: 10.1016/s0024-3841(00)00007-3.
Ladefoged, Peter, Jenny Ladefoged, Alice Turk, Kevin Hind & St. John Skilton. 1998. Phonetic structures of Scottish Gaelic. *Journal of the International Phonetic Association* 28. 1–41. DOI: 10.1017/s0025100300006228.
Le Dû, Jean. 1986. A sandhi survey of the Breton language. In Henning Andersen (ed.), *Sandhi phenomena in the languages of Europe*. Berlin, New York, Amsterdam: Mouton de Gruyter.
Lewin, Christopher. 2022. Continuity and hybridity in language revival: The case of Manx. *Language in Society* 51. 663–691. DOI: 10.1017/S0047404521000580.
Lieber, Rochelle. 1992. *An integrated theory of autosegmental processes*. Albany: SUNY Press.
Llwyd, Alan. 2007. *Anghenion y gynghanedd*. Swansea: Cyhoeddiadau Barddas.
Lowe, Jenefer. 2018. Language revitalization in Cornwall. *The Routledge Handbook of Language Revitalization*.
Morris, Jonathan. 2021. Social influences on phonological transfer: /r/ variation in the repertoire of Welsh-English bilinguals. *Languages* 6. DOI: 10 . 3390 / languages6020097.
Morris-Jones, John. 1913. *A Welsh grammar: Historical and comparative*. Oxford: Clarendon.
Morris-Jones, John. 1980. *Cerdd dafod*. Cardiff: Gwasg Prifysgol Cymru.
Nance, Claire. 2015. 'New' Scottish Gaelic speakers in Glasgow: A phonetic study of language revitalisation. *Language in Society* 44. 553–579. DOI: 10 . 1017 / s0047404515000408.
Ní Chiosáin, Máire. 1991. *Topics in the phonology of Irish*. University of Massachusetts Amherst. (Doctoral dissertation).

Ní Chiosáin, Máire. 1999. Syllables and phonotactics in Irish. In Harry van der Hulst & Nancy A. Ritter (eds.), *The syllable: Views and facts*, 551–575. Berlin & New York: Mouton de Gruyter.

Ní Chiosáin, Máire, Pauline Welby & Robert Espesser. 2012. Is the syllabification of Irish a typological exception? An experimental study. *Speech Communication* 54. 68–91. DOI: 10.1016/j.specom.2011.07.002.

Ó Maolalaigh, Roibeard. 1995. The development of eclipsis in Gaelic. *Scottish Language* 14–15. 158–173.

Ó Siadhail, Mícheál. 1989. *Modern Irish: Grammatical structure and dialectal variation*. Cambridge: Cambridge University Press.

Press, Ian. 2011. *A grammar of modern Breton*. Mouton: Walter de Gruyter.

Pyatt, Elizabeth J. 1997. *An integrated model of the syntax and phonology of Celtic mutation*. Cambridge: Harvard. (Doctoral dissertation).

Rogers, Henry. 1972. The initial mutations in modern Scots Gaelic. *Studia Celtica* 7. 63–85.

Smith, Norval. 1999. A preliminary account of some aspects of Leurbost Gaelic syllable structure. In Harry van der Hulst & Nancy A. Ritter (eds.), *The syllable: Views and facts*, 577–630. Berlin: Mouton de Gruyter.

Stewart, Thomas W. 2004. *Mutation as morphology: Bases, stems, and shapes in Scottish Gaelic*. Ohio State University. (Doctoral dissertation).

Teare, Rob & Max Wheeler. 2017. *Is there "homorganic inhibition" in Classical Manx? A corpus investigation*. https://www.academia.edu/35241836/Is_there_homorganic_inhibition_in_Classical_Manx_A_corpus_investigation (5 June, 2025).

Thomas, Alan R. (ed.). 2000. *The Welsh dialect survey*. Cardiff: University of Wales Press.

Williams, Briony. 1989. *Stress in modern Welsh*. Bloomington: IULC.

Chapter 7

Hebrides English in the 1980s and now

Ian Clayton[a] & Cynthia Shuken[b]

[a]University of Nevada, Reno [b]Independent scholar

Hebrides English is a distinct Scottish Gaelic-influenced variety of English spoken in the Hebridean island chain. The variety is most likely of relatively recent origin: before the second half of the nineteenth century, English was not widely spoken in this part of Scotland, where until recently Gaelic has been the dominant language. Hebrides English exhibits a variety of distinctive linguistic features which may be due to widespread bilingualism among its speakers in both Gaelic and English, including preaspiration and sibilant devoicing. The present study examines these two features, as well as glottalization, a possible incoming feature, in previously unanalyzed archival material from the early 1980s collected on the islands of Lewis, Harris, and Skye. The chapter thus offers a glimpse of Hebrides English as it existed four decades ago, and taken in context with more recent research (e.g. Clayton 2017, 2018) allows further insight into a little-studied variety of English evidently in considerable flux.

1 Introduction

1.1 Outline of this study

This chapter has two chief goals. The first is to offer some historical perspective on the phonological characteristics of Hebrides English, based on archival data collected in the early 1980s, and in so doing help to develop a picture of Hebrides English as it was spoken forty years ago. The second goal is to point to certain changes that have taken place in Hebrides English in the intervening decades.

Ian Clayton & Cynthia Shuken. 2025. Hebrides English in the 1980s and now. In Andrew Carnie, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond & Luis Irizarry (eds.), *Foundational approaches to Celtic linguistics*, 247–281. Berlin: Language Science Press. DOI: 10.5281/zenodo.15654869

The chapter makes use of two main sources of data. The first is the large body of sociolinguistic interviews conducted by Cynthia Shuken in the early 1980s.[1] Much of the data from that collection of material has never previously been analyzed or published (Shuken 1984, 1985, and 1986 offer analyses of subsets of the material). Shuken recorded interviews with nearly 100 individuals from Lewis, Harris, and Skye. Upon her departure from academia in the late 1980s, she placed these recordings and associated questionnaires in an archive at the University of Edinburgh School of Scottish Studies, where they have remained effectively untouched since. This collection represents an invaluable and unique record of Hebrides English as it was used in that era and region, and contains ample material to support a much more wide-ranging analysis than current space permits.

The second body of data referred to in this chapter is that collected by the first author during fieldwork in 2013 (cf. Clayton 2015, 2017, 2018). These studies included data from 24 speakers of Hebrides English, representing nine islands: Skye, Harris, Lewis, North Uist, South Uist, Grimsay, Raasay, Tiree, and Vatersay. While the two collections are not perfectly parallel, differing in numbers of participants, geographic scope, and to some extent in methodology, there is sufficient overlap between them to permit a useful examination of the trajectory of two sociophonetic variables over the time period in question: preaspiration and glottalization. The present chapter also considers sibilant devoicing as it occurs in Shuken's data; the Clayton study did not address such devoicing and so allows no comparative analysis of this feature.

The balance of the chapter proceeds as follows: Section 2 describes a number of distinguishing characteristics of Hebrides English, and provides a review of the existing literature on the topic. Section 3 describes the material in the Shuken archive in detail. Section 4 presents the analyses of preaspiration, glottalization, and devoicing, while Section 5 offers a summary of the results and a discussion.

2 Background and prior research

2.1 What is Hebrides English?

The remarkably diverse linguistic landscape found in Scotland is in large part due to the prolonged conditions of linguistic contact that have obtained in the region

[1]While the first author is the writer of this chapter and conducted the bulk of the analysis it incorporates, the chapter depends crucially on Cynthia Shuken's very substantial prior work, and indeed was possible only because she made her collected materials available to the first author. Upon request, Dr. Shuken agreed to be identified as a co-author, despite no longer working in academia. Any errors the chapter contains are solely those of the first author.

for centuries. The island groups of Shetland and Orkney to the northeast of the Scottish mainland, for example, are home to a variety of insular Scots influenced by Norn, the North Germanic language once spoken in those islands (Melchers 2008, McClure 1994, Barnes 1984, Wells 1982, Catford 1957a). Meanwhile, in the Highland and Island regions of northwestern Scotland one encounters English varieties influenced to a greater or lesser degree by the lexicon, syntax, and phonology of Scottish Gaelic, the Celtic language that has been, until the last few decades, the dominant language in the region (McClure 1994, Shuken 1984, 1985, Clement 1997, Stuart-Smith 2008, Clayton 2017, 2018).

While these Gaelic-influenced varieties are sometimes referred to collectively as *Highland and Island English*, they are by no means uniform, and further subdivisions are possible. Frequently, for example, a distinction is made between Highland English on the one hand, and Hebrides (or Hebridean or Island) English on the other (Clement 1980, 1997, Shuken 1985, Sabban 1985, Bird 1997). The basis of this division is the degree and kind of presumed contact with Gaelic. Thus by *Highland English* is usually meant the form spoken in the mainland Highland regions more or less adjacent to Scots-speaking areas. In this part of the Highlands, Gaelic has not been widely spoken for many decades if not longer, so linguistic contact between Gaelic and English is more a matter of historical record than contemporary fact. Hence, such characterizations of Highland English as "standard English on a Gaelic substratum" (Mather & Speitel 1986), or as "post-Gaelic English" (Catford 1957b) are reasonable ones.

By contrast, *Hebrides English* refers self-evidently to the variety of Scottish English frequently heard in the Hebridean island chain, especially those islands where Gaelic has at least until very recently remained a widely used language, and where bilingualism is commonplace if not universal. Though (Catford 1957b's) description of Hebrides English (HE hereafter) as "the English spoken, with varying degrees of imperfection, as a second language, by persons whose mother tongue is still Gaelic" is difficult to sustain six decades on as usage of Gaelic in the area continues to decline (Ó Giollagáin et al. 2020), it is nevertheless true that the degree of contact between English and Gaelic is much closer in these islands than anywhere else in Scotland.

Another factor likely to be relevant is the manner in which influences from English, or from Scots, have made their way into formerly Gaelic-dominant communities. Clement's (1980) point that learning English at school is quite distinct from learning Scots from one's neighbors remains a perspicuous one. English was largely introduced to the islands and more remote parts of the Highlands via instruction in the schools, upon English being made compulsory by the Education Act of 1872 (Withers 1984). By the 1970s, bilingualism in English was

effectively universal among Gaelic speakers, though even today one can readily find native Gaelic speakers who knew little or no English before beginning primary school. Thus, it is probably appropriate to regard Hebrides English as a variety of schoolroom Scottish Standard English that features significant Gaelic influences.

This description of HE, however, does not account for recent influences from urban varieties of Scots or Scottish English. The larger Island communities, such as Portree on Skye, and Stornoway on Lewis are prominent commercial centers with significant numbers of English- and Scots-speaking incomers with concomitant influences on local language use (Shuken 1984, Speitel 1981, Clement 1980). These influences seem to include the introduction of velarized [ɬ] (Shuken 1985) and glottal reinforcement (Clayton 2018, Wells 1982, Trudgill 1988, Stuart-Smith 1999a,b); there may well be others.

Finally, it must be stressed that English as used in the Hebrides is hardly uniform, as this chapter and other papers illustrate. What research has been published, largely focuses on the northern islands of the Hebrides, especially Skye, Harris, North Uist, and Lewis, and therefore presents a very incomplete picture. There remains a marked absence of research regarding English as it is spoken on the other Hebridean islands.

2.2 Possible Gaelic influences in Hebrides English

Discussions of HE in the literature frequently mention certain features that may well be due to contact with Gaelic. They include:

2.2.1 Morphosyntax

1. The "BE after V-ing" construction (Sabban 1985):

   (1) Tha mi dìreach air mo nigheadaireachd a' dhèanamh
   be.PRS I just PERF my washing PROG doing
   'I'm just after finishing my washing.'

2. Absence of the English present perfect construction (Sabban 1985):

   (2) Tha mi air a bhith nam bhanntrach-fhir a-nis airson sia bliadhna
   be.PRS I PERF ART be.VN in-my widower now for six years
   'I'm a widower now for six years.'

3. Cleft constructions (Shuken 1984):

   (3) ‘S e Gàidhlig a bhruidhinn sinn an còmhnaidh aig an
   cop.PRS it Gaelic REL spoke we always at the
   taigh
   home
   ‘It was always Gaelic we spoke in the home.’

2.2.2 Phonology

1. There is no voicing contrast among Gaelic sibilants. HE realizations like *measure* as [mɛʃur], and *just* as [tʃʌst] are thus presumably due to Gaelic influence (Shuken 1985, Wells 1982).

2. In the Hebridean varieties of Gaelic, the unaspirated voiceless stops /p t k/ contrast with the voiceless aspirated series /pʰ tʰ kʰ/. There are no voiced stops. In like vein, many HE speakers omit voicing in English /b d g/, e.g. *glory* [klɔɾi], *ragged* [rækɪt], *building* [pɪltɪŋ] (Grant 1913, Borgstrøm 1940, Oftedal 1956, Wells 1982, Ladefoged et al. 1998, Gillies 2009).

3. In many varieties of Gaelic, the aspirated stop phonemes /pʰ tʰ kʰ/ are preaspirated in non-initial positions, for instance as [ʰp ʰt ʰk] (the typical Lewis pattern), or as [ʰp ʰt ˣk] (as in Skye and Harris). In similar fashion, voiceless stops in HE may feature preaspiration: *top* [tʰɔʰp], *not* [nɔʰt], *cat* [kʰæʰt], *bake* [beʰk] (Borgstrøm 1940, Oftedal 1956, Shuken 1979, Ní Chasaide 1985, Ladefoged et al. 1998, Clayton 2010, Nance & Stuart-Smith 2013). Lewis speakers have been particularly singled out in this regard, e.g. Borgstrøm (1940) notes that “postvocalic *k*, *t*, *p* are usually rendered by Lewismen as ^{h}k, ^{h}t, ^{h}p, and nobody seems to be aware that this is not the proper English pronunciation” (p. 21). Likewise, Shuken (1984, 1985) found abundant preaspiration in Lewis English, but rather less among Skye and Harris speakers.

4. Epenthetic vowels: In Gaelic, epenthesis affects post-vocalic underlying VC_1C_2 sequences: the vowel must be short, C_1 must be a nasal or liquid, and C_2 must be a non-homorganic consonant apart from /pʰ tʰ kʰ/. In such instances, an epenthetic vowel will be inserted between the two consonants and which matches the vowel preceding the CC cluster, e.g. *arm* /arm/ ‘army’ becomes [a.ram], *airgead* /ɛrʲkʲət/ ‘money’ becomes [ɛrʲɛkʲət]. In HE, a similar, though not identical, pattern obtains among some speakers, though it seems to apply in slightly different phonological conditions:

an epenthetic vowel may appear in VC_1C_2 sequences in which C_1 is a nasal or liquid, just as in Gaelic, though in HE the CC may be homorganic: *milk* [mɪlɪk], *farm* [faram], *kiln* [kɪlɪn] (Hammond et al. 2014, Smith 1999, Bosch & de Jong 1997).

5. Dental stops: the coronals /t d/ may be dentalized as in Gaelic, e.g. *tree* [t̪ʰɾi] (Grant 1913, Holmer 1957, Dorian 1978, Holmer 1981, Shuken 1984, Ladefoged et al. 1998).

6. Retroflex /s/: HE frequently features retroflexion of /s/ after /r/, as in *first* [fɻ̩st], *horse* [hɔɻ̩ʂ], *parcel* [paɻ̩səl], a pattern also found in Gaelic, e.g. *spors* [spoːɹʂ] 'sport', *dearrsach* [tʲaɹʂax] 'bright' (Borgstrøm 1940, Wells 1982, Shuken 1984, Stuart-Smith 2008).

7. Laterals: The lateral liquid common among HE speakers is a clear [l]. Elsewhere in Scotland, a velarized [ɫ] frequently predominates. While some varieties of Gaelic do have both velarized and non-velarized laterals, the velarized variant is rare in the northern Hebrides (Holmer 1957, 1981, Ternes 1973, Dorian 1978, Wells 1982, Shuken 1984, 1985, Ó Murchú 1989, Johnston 1997, Ladefoged et al. 1998, Stuart-Smith 1999a, Nance 2014).

2.3 Prior research

2.3.1 Published studies of Hebrides English

Published in-depth studies of Hebrides English are notable for their scarcity. Those that are available point both to the distinctive character of the variety (in particular those features potentially of Gaelic origin, as described above), and to progressive convergence with mainland varieties through a process of dialect leveling, or what Hickey (2013) refers to supraregionalisation: a process in which a particular language variety loses those identifiably local features that have previously set it apart from a more broadly distributed (and perhaps less stigmatized) variety. Below I describe the work of Sabban (1985), Shuken (1984, 1985, 1986), and Clayton (2015, 2017, 2018). I delve into the Clayton and Shuken studies in fair detail because they are the most relevant to the present chapter, particularly Shuken's work.

2.3.2 Sabban (1985)

Sabban (1985) is a morphosyntactic study of the English spoken in the Hebridean islands of Skye and North Uist. It is to date the only published study of HE morphosyntax, so far as I am aware. Sabban examined two patterns: the *BE after*

V-ing construction, and the absence of the present perfect (cf. Section 2.2). Participants included 29 bilingual speakers, aged 6 to 89. Sabban found that these two patterns were overwhelmingly concentrated in speakers over 50, while speakers under 20 used them very rarely if at all. These results strongly suggest that these morphosyntactic features were obsolescent in these areas at the time of her study, a situation that is likely only to have advanced in the intervening four decades.

2.3.3 Shuken (1984, 1985, 1986)

In the late 70s and early 80s, Cynthia Shuken conducted the first systematic sociophonetic study of Hebrides English (1984, 1985, 1986), with the aims of providing a thorough phonetic description of the variety, and identifying any key sociolinguistic variables that might be functioning as indices of identification with socioeconomic groups. She recorded nearly one hundred Hebridean English speakers, representing the islands of Harris, Lewis, and Skye. Her material therefore is confined to the English spoken in these three islands. The participant pool was balanced between male and female speakers, and grouped into three age brackets (17–25, 35–45, and 55+). In addition, participants included rural speakers from all three islands, and urban speakers drawn from Stornoway, the chief town on Lewis; and from Portree, the principal population center on Skye. The Skye and Lewis urban speakers were further subdivided into English monolinguals and English-Gaelic bilinguals. Since no comparable urban center exists on Harris (its largest community, Tarbert, had a population of 464 in the 1981 census), Shuken's 14 Harris speakers were all categorized as rural bilinguals. At the time of the study, the rural areas of all three islands were overwhelmingly bilingual, and so she made no effort to include rural monolinguals. Her collected material included recordings of wordlists, reading passages, and unstructured conversation.

Three publications resulted from this study. The first, Shuken (1984), provides a brief history of the variety as well as a profile of the phonology and morphosyntax of HE. No quantitative analysis is presented in this paper.

The second publication, Shuken (1985), was a sociophonetic analysis of HE based on data from the youngest age bracket in her participant group, i.e. 17–25 year olds, including rural bilinguals, urban bilinguals, and urban monolinguals. This paper compared the relative abundance of two HE variables of potential Gaelic provenance, the unvelarized or clear [l] and the preaspiration of voiceless stops, with two variables of likely external origin, namely the velarized lateral [ɬ]

and the glottalization of voiceless stops (both in the form of glottal replacement or T-glottaling, and glottal reinforcement).

Shuken found that the use of velarized [ɬ] rather than clear [l] was much more abundant among urban speakers compared to rural speakers in both Skye and Lewis, much more abundant on Skye overall than Harris or Lewis, and most abundant among urban monolinguals from Skye. Conversely, clear [l] dominated among rural speakers, especially those from Lewis.

Shuken found that preaspiration was more common among Lewis speakers than those from Harris or Skye, that it was slightly more common among rural speakers than urban speakers in Lewis, and that it was of similar frequency in both rural and urban speakers in Skye. On the other hand, she found that glottalization was much more abundant among Skye speakers than those from Harris or Lewis, and more abundant among urban speakers than rural speakers in both Lewis and Skye. Finally, she found that glottal replacement was more common among Skye speakers than those from Harris or Lewis and more common among urban speakers than rural speakers.

Given the greater relative abundance of velarized [ɬ] and glottalized forms among young speakers in urban communities, and among those in greater proximity to the mainland, Shuken concluded that there was evidence for a significant and growing degree of mainland influence on HE at the time of this study.

Shuken (1986) examined the phenomenon of vowel lengthening among HE speakers in relation to the patterns seen in Gaelic, in Scottish Standard English, and in Received Pronunciation. She found that a small number of word pairs like *great ~ grate* and *maid ~ made* were in contrast among some speakers, such that one member of the pair reliably differed from the other in vowel length. Shuken suggested that the feature is potentially due to influence from Gaelic, where contrastive vowel length is pervasive, e.g. *srac* [s̪t̪raxk] 'rend!' versus *stràc* [s̪t̪ra:xk] 'stroke'. However, in HE such unconditioned contrasts were seen to be sporadic, and with no consistent pattern between speakers.

Apart from the few instances of contrastive lengthening, Shuken found two patterns of conditioned vowel length which varied in frequency by island. Speakers from Harris and Lewis were most likely to follow the pattern widely observed in Scottish Standard English, wherein (most) vowels are predictably lengthened before voiced fricatives, /r/, and morpheme boundaries (Aitken 1981). Meanwhile, Skye speakers of HE were more likely to follow the Received Pronunciation pattern of lengthening vowels predictably before voiced consonants and in open syllables. In other words, the RP pattern had a greater degree of influence in Skye, the island closest to the mainland, than in either of the two more remote islands.

In general, then, Shuken's published work on HE points to a linguistic situation in considerable flux, with evidence of expanding influence from mainland varieties of English.

2.3.4 Clayton (2015, 2017, 2018)

The only work on HE to be published since the mid-1980s is found in Clayton (2015, 2017, 2018). This study surveyed 24 bilingual speakers of HE and Gaelic, 10 male and 14 female. A range of islands within the Hebrides archipelago were represented, though unfortunately it was extremely difficult to identify more than one or two speakers from some of the islands. These islands included Skye (8 speakers), Harris (1), Lewis (8), North Uist (2), South Uist (1), Grimsay (1), Raasay (1), Tiree (1), and Vatersay (1). The study addressed the incidence of preaspiration and of glottalization (both as glottal reinforcement and as T-glottaling). Preaspiration was found to be abundant only among a subset of older female speakers from Lewis, suggesting both that the feature is concentrated in the English of Lewis (cf. Borgstrøm 1940, Shuken 1984, Sabban 1985), and that preaspiration is falling out of use among younger speakers. Glottal reinforcement was most abundant in younger speakers, particularly younger female speakers, though no clear pattern was found indicating that these features were comparatively more abundant in Skye than in Harris or Lewis (cf. Shuken 1985 with a similar result). Finally, T-glottaling was highly restricted in its occurrence: it only occurred in the word *rotten*, and predominantly among Skye speakers. Neither age nor sex appeared to affect the frequency of T-glottaling. The greater abundance among younger speakers suggests that glottal reinforcement, well-established in urban varieties of Scottish English (Stuart-Smith 2008, 2003, 1999a,b, Wells 1982), are advancing in Hebrides English as well.

3 The Shuken archive

3.1 Overview

Cynthia Shuken's collected materials are archived at University of Edinburgh's School of Scottish Studies. They include about 85 tape reels, of which 40 contain sociolinguistic interviews relevant to HE; detailed IPA transcriptions of some of the recorded interviews; 92 completed sociolinguistic questionnaires; and an assortment of fieldwork notes.

The tapes in the archive seem to be copies in most instances rather than originals. Often, material from one speaker is scattered across several tape-reels. In a

small number of cases, identical material occurs on more than one tape. While the tape-reels are in labeled boxes, in most cases the labels correspond only partially, or else not at all, with the material actually contained on the tape. These factors presented significant organizational challenges which had to be overcome before the present analysis could commence.

All tape-reels in the collection were digitized by the very helpful archivists at the first author's request; these digitized versions were used for all the new analysis presented in this chapter.

3.2 Recording procedures

While a small number of Shuken's participants were recorded in a laboratory setting at the University of Edinburgh, most were recorded in the field using a portable Uher reel-to-reel tape recorder and a table-top Sennheiser microphone. Recording locations varied widely, and included the participants' homes, workplaces, and local primary schools among other settings. Most recordings were made in quiet conditions, but given the nature of fieldwork, some recordings inevitably feature a degree of ambient noise, such as doors slamming, background conversations, vehicle engines idling, and so forth. In no case, however, did such background noises significantly impede the type of analysis carried out for this chapter. The exact model and specifications of the recorder and microphone are unavailable at this point, but it seems safe to assume that they were industry standard for the era. The overall audio quality of the recorded material is generally quite high.

During a typical recording session, Shuken would first ask the participant to state their name and place of residence. She would then have them read two passages in English. Gaelic-speaking participants would then read a list of Gaelic words, and then move on to the English word list. All words were given to participants as printed lists without using a frame sentence; all participants encountered identical lists. Usually, Shuken would conclude each session by having participants complete a sociolinguistic questionnaire; frequently, the recording includes the participant's commentary on the questionnaire, though not always.

3.3 Questionnaires

Shuken used a detailed 7-page sociolinguistic questionnaire to collect demographic information about her participants as well as their cultural and political attitudes. Two versions of the questionnaire were prepared: one for the use of Gaelic-speaking bilinguals, and one for English-speaking monolinguals. Among

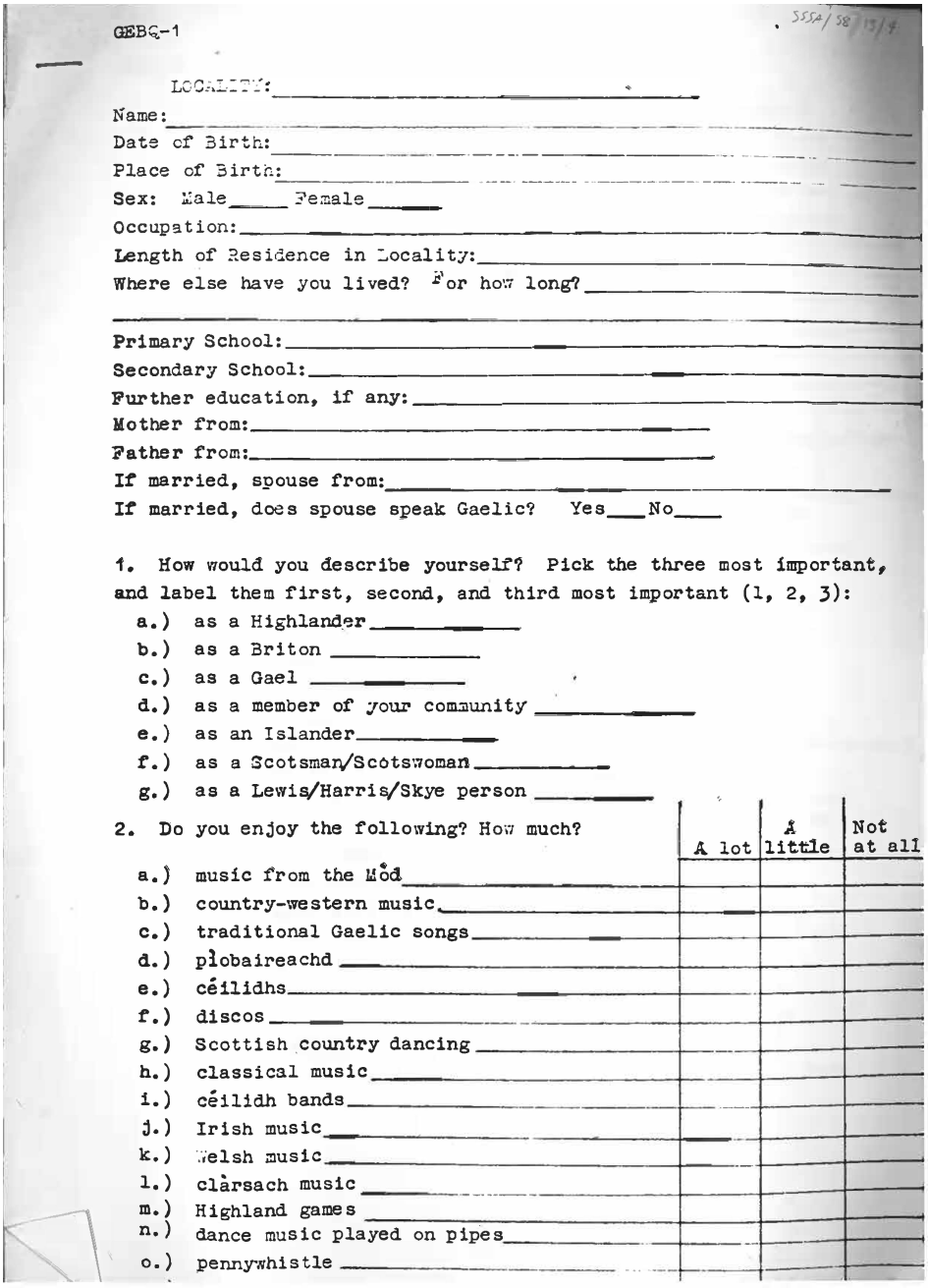
other things, the questionnaire asked about each participant's place of birth and that of their parents; their residence history, education and employment; their degree of personal identification with the local community and region, with Scotland, and with the United Kingdom; their preferences and degree of knowledge of cultural matters such as traditional Highland music, literature, and history; the frequency with which they spoke Gaelic (if bilingual) and with whom; their perceptions of the relative attractiveness and correctness of various dialects of Gaelic and English; and their degree of pride in speaking a distinct variety of English. The first page of the bilingual version of the questionnaire is reproduced in Figure 1 (page 258).

The archive contains completed sociolinguistic questionnaires for 92 individuals: 46 from Lewis, 14 from Harris, and 32 from Skye. While the questionnaires are numbered consecutively (L1-L50 for Lewis, H1-H14 for Harris, and S1-S50 for Skye), there are gaps for 4 questionnaires from the Lewis group and 18 from the Skye group, suggesting that the corresponding recordings were not made. Recorded interviews corresponding to almost all of the completed questionnaires could be located on the tape reels, with the exception of three participants from Lewis, one from Harris, and three from Skye; however, Shuken's notes include complete, highly detailed transcriptions of the English wordlist and the two reading passages for one of the three missing Skye speakers (S12), and for the Harris speaker (H6). Thus, there is audio available for 85 interviews plus detailed transcripts of two more interviews.

3.4 Participants

Shuken's full participant pool includes 92 speakers of HE from the islands of Lewis, Harris, and Skye, as represented in the archive (if we include the seven participants for whom we have questionnaires but no recordings). The pool was designed to be evenly balanced for sex and was stratified by age group (17–25, 35–45, 55+). In addition, Shuken included rural speakers from all three islands as well as urban speakers drawn from Stornoway, the chief town on Lewis, and from Portree, the principal population center on Skye. The urban speakers included both monolingual HE speakers and bilingual speakers of HE and Gaelic.

A bit more could be said at this point about the towns of Stornoway (on Lewis) and Portree (on Skye). At the time of Shuken's study, the parish of Stornoway had a population of approximately 7900, according to the 1981 census (Shuken 1985). The town's economy was, and still is, centered on retail, social services, tourism, textiles, aquaculture, and offshore gas and oil extraction (Highlands &



GEBQ-1

LOCALITY: _____

Name: _____

Date of Birth: _____

Place of Birth: _____

Sex: Male_____ Female_____

Occupation: _____

Length of Residence in Locality: _____

Where else have you lived? For how long? _____

Primary School: _____

Secondary School: _____

Further education, if any: _____

Mother from: _____

Father from: _____

If married, spouse from: _____

If married, does spouse speak Gaelic? Yes___ No___

1. How would you describe yourself? Pick the three most important, and label them first, second, and third most important (1, 2, 3):

a.) as a Highlander _____
b.) as a Briton _____
c.) as a Gael _____
d.) as a member of your community _____
e.) as an Islander _____
f.) as a Scotsman/Scotswoman _____
g.) as a Lewis/Harris/Skye person _____

2. Do you enjoy the following? How much?	A lot	A little	Not at all
a.) music from the Mòd			
b.) country-western music			
c.) traditional Gaelic songs			
d.) pìobaireachd			
e.) céilidhs			
f.) discos			
g.) Scottish country dancing			
h.) classical music			
i.) céilidh bands			
j.) Irish music			
k.) Welsh music			
l.) clàrsach music			
m.) Highland games			
n.) dance music played on pipes			
o.) pennywhistle			

Figure 1: Page from Shuken's Gaelic-English bilingual questionnaire

Enterprise 2020). Stornoway's population has declined somewhat since 1981; the estimated population as of 2020 was 7280 (National Records of Scotland 2023).

Like Stornoway, Portree is a regional commercial center, though more heavily dependent on tourism and arguably more subject to outside influence as a consequence. The completion of a road bridge connecting Skye to the mainland in 1995 has only strengthened the commercial and social ties between Skye and the mainland since Shuken conducted her fieldwork (Highlands & Enterprise 2007) and thus magnified the potential for mainland linguistic influence. The 1981 census recorded a population of 1414 in Portree compared to an estimated 2,310 in 2020 (National Records of Scotland 2023), indicating a significant pattern of growth in the community.

Shuken was able to record participants representing nearly every possible combination of demographic factors in her planned speaker population. For most combinations of the factors of island, age, rural/urban, and monolingual/bilingual, her recorded interviews include at least one male and one female speaker. The majority achieved her target of at least two male and two female speakers for each of these conditions. Where unfilled gaps do occur, these include female urban monolinguals from Lewis in the 55–65 age group, male and female rural bilinguals from Skye in the 17–25 age group, Skye rural bilingual males in the 35–45 age group, and Skye urban monolingual males in the 55–65 age group; no speakers representing these subgroups are present in the recorded material. Conversely, in a small number of conditions "surplus" speakers were recorded, i.e. more than the target of two women and two men per condition. For the present study, the data from only two male and two female speakers per condition were included in the analysis, in order to prevent overrepresenting any subpopulation.

The final speaker pool for the present study contains 70 individuals, with the distribution shown in Table 1.

3.5 Stimuli and reading passages

Shuken's full stimulus set consists of 319 English words.[2] While the majority of the stimuli presented no difficulties for participants, the recordings do indicate that several of the words in the list were unfamiliar to many people, who struggled to read them accurately as a consequence. Problematic words included *quoit, adroit, gesture, loathe,* and *rouge.* These words were excluded from the present analysis. In addition, the word *use* was excluded because it was orthographically ambiguous: participants were unsure whether the verb (with a final /z/)

[2]The full set of stimuli is available upon request to the first author.

Table 1: Distribution of participants according to island (Lewis, Skye, Harris), age group, sex, and urban/rural, bilingual/monolingual (e.g. LRB = Lewis, rural, bilingual)

Age/sex	LRB	LUB	LUM	SRB	SUB	SUM	HRB
17–25 F	L1 L2	L13 L14	L25 L26	-	S1	S25 S26	H1 H2
17–25 M	L3 L4	L15	L27 L28	-	S3 S15	S27 S28	H3 H4
35–45 F	L5 L6	L17	L29 L30	S37 S38	S5 S6	S48	H5 H6
35–45 M	L7 L8	L19 L20	L31 L32	-	S7 S19	S31 S49	H7 H8
55+ F	L9 L10	L21 L22	-	S9 S10	S40 S42	S33 S34	H9 H10
55+ M	L11 L12	L23 L24	L35 L36	S11 S12	S41 S50	-	H11 H12

was meant, or the noun (with a final /s/), and so the word could not be reliably analyzed.

In addition to the list of individual words, Shuken also presented participants with two short reading passages. Analysis of the data offered by these reading passages is ongoing, and is not included in this chapter. For the three variables considered in the present analysis, the relevant subsets of stimuli are illustrated in Tables 2 and 3.

Table 2: Stimuli for preaspiration and glottalization of /p t k/

/p/	/t/				/k/	
chip	*about*	*boot*	*grate*	*soot*	*actually*	*plucking*
cup	*bait*	*booty*	*great*	*suit*	*choke*	*spoke (n.)*
dip	*beat*	*caught*	*idiot*	*water*	*creak*	*spoke (v.)*
loop	*beauty*	*coat*	*illuminate*		*creek*	*tack*
ship	*beet*	*cot*	*mate*		*hawk*	*taxes*
staple	*bet*	*crate*	*petal*		*joke*	*taxis*
stoop	*bit*	*cut*	*put*		*leak*	*weak*
stripes	*bite*	*feature*	*quite*		*leek*	*week*
tip	*boat*	*foot*	*right*		*lock*	

Table 3: Stimuli used to evaluate voicing contrasts in sibilants

final /z/		initial /z/	final /dʒ/	medial /dʒ/	medial /ʒ/
boys	*rays*	*zeal*	*badge*	*dredges*	*measure*
cause	*resume*		*cadge*		
caws	*rose*				
days	*rouse*				
daze	*rows (boat)*				
enthuse	*rows (fights)*				
ewes	*seas*				
fez	*seize*				
fuzz	*sighs*				
halves	*size*				
noise	*stretches*				
ooze	*ways*				
presume	*weighs*				
raise					

4 Analysis

4.1 Analysis of preaspiration

In Gaelic, the realization of preaspiration tends to vary depending both on the dialect in question and the affected consonant. In many dialects, particularly those of the eastern mainland now moribund or extinct, preaspiration does not occur at all. Among the varieties of Gaelic found in Lewis, Harris, and Skye, there are two main preaspiration patterns. On the one hand is the Lewis pattern, in which preaspiration is typically realized as a period of glottal frication or breathy voice [VhC ~ VɦC] occurring between a vowel and a following voiceless stop. On the other hand, we have the pattern found in Harris and Skye, where before labial and coronal stops preaspiration appears much as in Lewis, i.e. as glottal frication or breathy voice, but before dorsals it is typically realized as homorganic oral frication, i.e. [ç͡c x͡k] (Nance & Stuart-Smith 2013, Clayton 2010, Ladefoged et al. 1998, Ó Dochartaigh 1994-1997, Ní Chasaide 1985, Borgstrøm 1940). The oral form [x͡p x͡t x͡k] alone is typical of certain dialects, particularly in Perthshire and Argyll (Iosad et al. 2015, Ó Murchú 1989).

In HE, by contrast, preaspiration seems to be rather less variable. There it typically takes the form of glottal frication or breathy voice (Figure 2). Homorganic

oral frication does infrequently occur, particularly with dorsals (Clayton 2018, 2017, 2015, Shuken 1985).

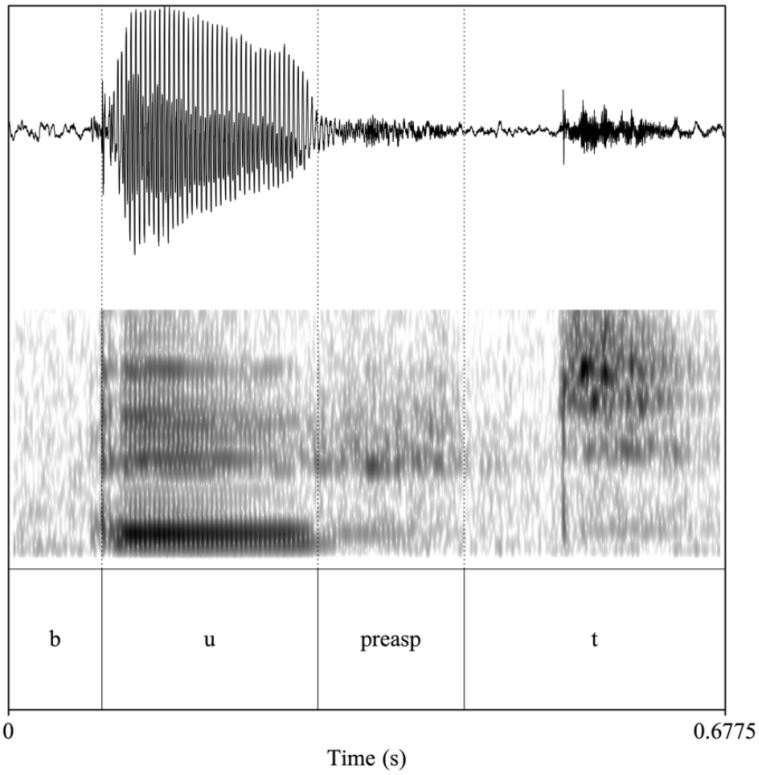


Figure 2: Preaspiration in the word *boot* as produced by L02, a female rural bilingual from Lewis in the 17–25 age group.

For the present study, tokens of preaspiration were identified through inspection of the waveform and spectrogram in Praat (Boersma & Weenink 2020). Of the 319 stimuli produced per speaker, 57 contained the voiceless stops /p t k/ postvocalically in medial or final position, and thus were potential contexts for preaspiration. Of these, 9 stimuli contained /p/, 31 contained /t/, and 17 contained /k/ (Table 2). Totaled across the 70 participants, there were thus 3990 relevant tokens. Of these, 658 tokens (16.5%) were identified as featuring some form of preaspiration (Table 4). Coronal tokens were by far the most numerous, numbering 452 (68.7%) of the total, followed by dorsals (152 tokens, 23.1%), then labials (54 tokens, 8.2%). No labial tokens at all were noted among the Harris and Skye speakers, and very few dorsals (4 out of 29, 13.8%) were present in those two groups. This distribution favoring coronals is similar to that found for HE in Clayton (2017, 2018),

but stands in sharp contrast to the pattern typically observed in Gaelic, where dorsals are favored in terms of frequency (Clayton 2010).

Table 4: Distribution of preaspirated tokens according to island, sex, age group, and place of articulation.

	Lewis		Harris		Skye		
	M	F	M	F	M	F	Total
Age 1 (17–25)							
p	7	13	0	0	0	0	20
t	50	79	1	5	0	0	135
k	15	27	0	0	0	0	42
Total	72	119	1	5	0	0	197
%	25.26	34.80	0.88	4.39	0.00	0.00	
Age 2 (35–45)							
p	6	10	0	0	0	0	16
t	57	81	1	4	2	1	146
k	23	26	0	0	1	0	50
Total	86	117	1	4	3	1	212
%	25.15	41.05	0.88	3.51	1.32	0.44	
Age 3 (55+)							
p	7	11	0	0	0	0	18
t	92	68	3	5	0	3	171
k	28	29	0	0	0	3	60
Total	127	108	3	5	0	6	249
%	37.13	47.37	2.63	4.39	0.00	1.50	
						Total:	658

We see a very uneven distribution of preaspiration according to social factors (Figures 3 and 4). Preaspiration was overwhelmingly localized to Lewis speakers, who produced 629 tokens of the total 658 observed (95.6%). Each of the Lewis speakers produced a minimum of 2 preaspirated tokens, with the exception of L20, an urban bilingual male in the 35–45 age bracket, who produced none.

Statistical analysis of the effects of demographic variables on preaspiration rate (Table 5) was conducted using a logistic mixed-effects model via the glmer function of the `lme4` package (Bates et al. 2015, Kuznetsova et al. 2017). Because of the very low rates of preaspiration in Skye and Harris speakers, these islands were excluded from statistical analysis. The final model included *word* as a random effect, and *sex*, *age*, and *rural/urban* as fixed effects. Age was included as

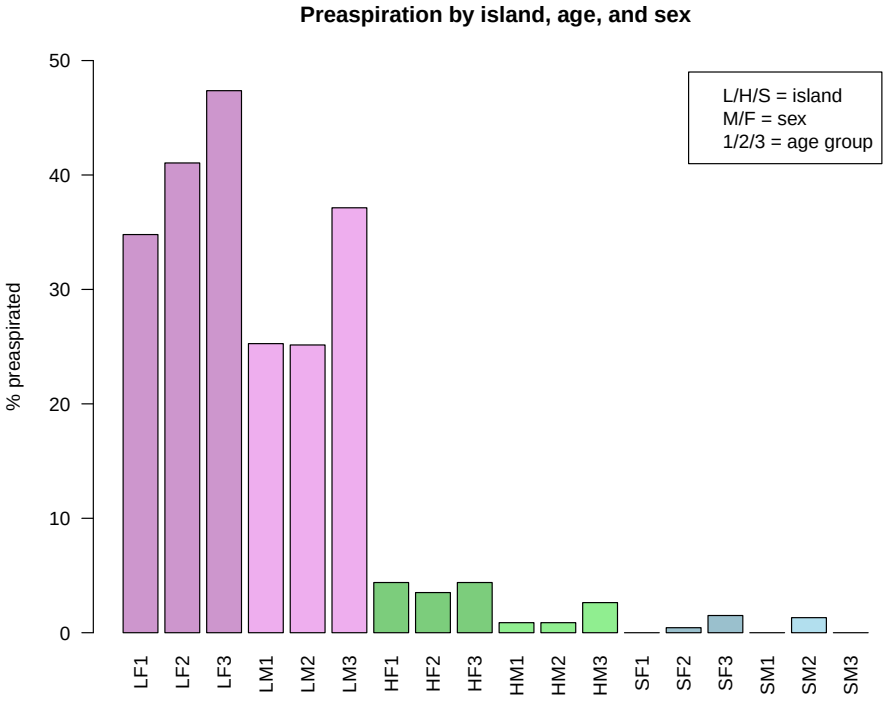


Figure 3: Rates of preaspiration according to island (Lewis, Harris, Skye), sex, and age group (1 = 17–25, 2 = 35–45, 3 = 55+)

a 3-way variable instead of continuous variable, to correspond with the three age bins among Shuken's participants; in addition, it was not always possible to identify the precise age of some speakers at the time of recording beyond the age bracket they occupied. While *speaker* is commonly included as a random effect in mixed-effect models (Baayen et al. 2008), its inclusion in the models used here tended to obscure the very variability the study hopes to highlight, and so *speaker* was not included as a factor in this and subsequent statistical models.

Among Lewis speakers, the frequency of preaspiration was skewed toward women overall (estimate $= -1.389$, $z = -4.039$, $p < 0.001$) (Figure 4c), and toward older speakers (Figure 4b). Older male speakers in particular preaspirated at significantly greater rates than the younger two groups of male speakers (estimate $= 1.568$, $z = 3.237$, $p < 0.001$). Urban speakers also preaspirated significantly less than rural speakers (estimate $= -1.869$, $z = -6.122$, $p < 0.001$).

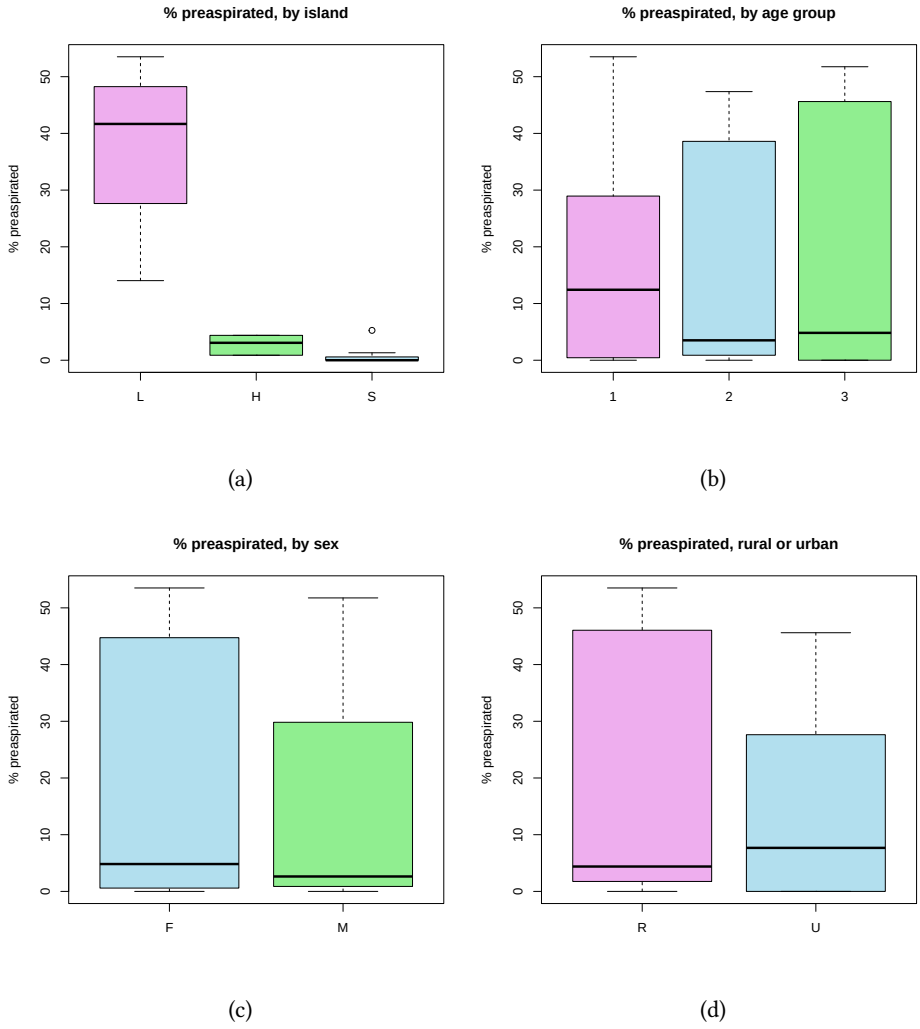


Figure 4: Comparisons of preaspiration rates by island (Lewis, Harris, Skye), age group (1 = 17–25, 2 = 35–45, 3 = 55+), sex, and rural vs. urban.

Table 5: Mixed-effects model of preaspiration rates among Lewis speakers. Intercept represents younger female rural speakers.

Random effects		Variance	SD	
word		4.349	2.085	
Fixed effects	Estimate	SE	z	p
(Intercept)	−0.030	0.373	−0.081	0.936
sexM	−1.389	0.344	−4.039	<0.001***
rurbU	−1.869	0.305	−6.122	<0.001***
age2	−0.591	0.340	−1.738	0.082
age3	−0.301	0.342	−0.880	0.379
sexM:rurbU	1.014	0.441	2.297	0.022*
sexM:age2	1.562	0.481	3.249	<0.001***
sexM:age3	1.568	0.484	3.237	<0.001***
rurbU:age2	1.471	0.430	3.422	<0.001***
rurbU:age3	1.636	0.454	3.607	<0.001***
sexM:rurbU:age2	−3.013	0.627	−4.808	<0.001***
sexM:rurbU:age3	−2.224	0.632	−3.519	<0.001***

4.2 Analysis of glottalization

Just as for preaspiration, glottalization is frequently associated with voiceless stops in HE. Effectively then the number of possible sites for glottalization in the current dataset is 3990, the same as for preaspiration. Of these 3990 tokens, 336 featured some degree of glottalization of the vowel just prior to the stop closure. Outright replacement of the consonant with a glottal stop (T-glottaling) was extremely rare in the Shuken data analyzed for this chapter and so not analyzed separately. Glottalization was identified in the data via irregularities in the timing of vocal fold pulsations (so-called "jitter"), or in the amplitude of the associated waveforms ("shimmer"). Fluctuations in the pitch track provided by Praat could also be observed in some glottalized tokens. Figure 5 illustrates glottalization in the word *leak* as produced by S25, an urban monolingual female speaker from Skye in the 17–25 age group.

Statistical analysis of the glottalization data (Table 7) was conducted by fitting a logistic mixed-effects model via the glmer function of the `lme4` package. The final model included *word* as random effect; and *age group*, *sex*, and *bilin-*

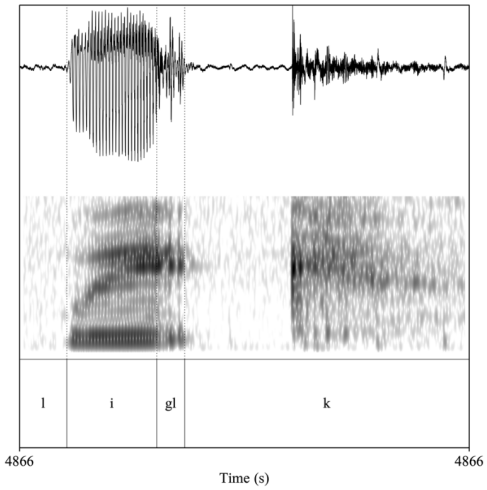


Figure 5: Glottalization in the word *leak* as produced by S25, a female monolingual from Skye in the 17–25 age group.

gual/monolingual as fixed effects. Interactions between the fixed effects were also modeled.

In terms of place of articulation, the distribution of glottalization is considerably more even than preaspiration: 191 of 2170 coronal tokens displayed glottalization (8.8%), 90 of 1190 dorsal tokens (7.6%), and 55 of 630 labial tokens (8.7%). No significant difference by place of articulation was found.

In terms of demographics, however, the distribution of glottalization was in certain respects the inverse of preaspiration (Figure 6). Skye speakers glottalized at far greater rates than Harris or Lewis speakers (Table 6; Figure 7a), glottalizing 251 observed tokens (74.7%), compared to 29 tokens (8.6%) among the Harris speakers and 56 (16.7%) among speakers from Lewis. Age was a significant factor (Figure 7b), where the older two brackets seem to have glottalized at proportionally higher rates, but the differences are very small (age 2: estimate $= 1.270$, $z = 3.357$, $p < 0.001$; age 3: estimate $= 0.922$, $z = 2.445$, $p = 0.014$). While women glottalized at slightly higher rates than men (Figure 7c), sex was not a statistically significant factor. Skye women did lead Skye men in the younger two age brackets, though glottalization was more common among older Skye men than women. Outside of Skye, glottalization could be described as abundant only among younger Lewis women; notably, 31 of the 35 tokens attributed to that group were produced by one individual, L26, a monolingual urban speaker. The other apparent cluster is among the older Harris women, though here all 11 tokens in that group were produced by one individual, H09. The other speaker in

the older female Harris group did not glottalize at all. Finally, monolinguals glottalized at significantly higher rates than bilinguals (Figure 7d; estimate = 3.628, $z = 9.975$, $p < 0.001$) at all age groups and in both male and female speakers.

Table 6: Distribution of glottalized tokens according to island, sex, age group, and place of articulation.

	Lewis		Harris		Skye		
	M	F	M	F	M	F	Total
Age 1 (17–25)							
p	2	6	0	1	3	7	19
t	1	22	0	5	18	40	86
k	1	7	0	1	9	17	35
Total	4	35	0	7	30	64	140
%	1.40	10.23	0.00	6.14	13.16	37.43	
Age 2 (35 -45)							
p	1	3	1	2	5	1	13
t	4	2	1	3	12	16	38
k	2	1	0	2	6	15	26
Total	7	6	2	7	23	32	77
%	25.15	2.11	0.88	6.14	1.32	0.44	
Age 3 (55+)							
p	0	0	0	2	12	9	23
t	2	2	2	9	29	23	67
k	0	0	0	0	17	12	29
Total	2	2	2	11	58	44	119
%	0.01	0.88	1.75	9.65	25.44	11.03	
						Total:	336

4.3 Analysis of devoicing

A commonly noted feature of HE is the absence among some speakers of voicing in obstruents, producing (near) mergers of /p t k/ ~ /b d g/, /tʃ/ ~ /dʒ/, and /s/ ~ /z/ (Wells 1982, Shuken 1984, Jones 2002, Stuart-Smith 2008). More rarely, a reversal of voicing is remarked; e.g., /s/ is realized as [z]. The usual assumption is that this trait has been acquired from the local varieties of Gaelic, which also lack a voicing contrast in such consonants, though contrast is maintained among stops via aspiration or preaspiration.

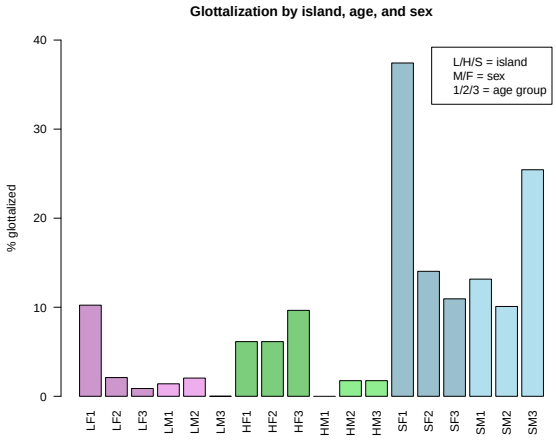


Figure 6: Comparison of glottalization rates by island (Lewis, Harris, Skye), sex, and age group (1 = 17–25, 2 = 35–45, 3 = 55+)

Table 7: Mixed-effects model of glottalization data. Intercept represents younger female bilingual speakers.

Random effects:		Variance	SD	
word		0.474	0.689	
Fixed effects:	Estimate	SE	z	p
(Intercept)	−3.971	0.349	−11.370	<0.001
sexM	−0.258	0.504	−0.513	0.608
age2	1.270	0.378	3.357	<0.001***
age3	0.922	0.377	2.445	0.014*
langM	3.628	0.364	9.975	<0.001***
sexM:age2	−1.270	0.640	−1.984	0.047*
sexM:age3	1.035	0.550	1.881	0.060
sexM:langM	−1.572	0.563	−2.794	0.005**
age2:langM	−3.791	0.511	−7.421	<0.001***
age3:langM	−2.088	0.468	−4.459	<0.001***
sexM:age2:langM	3.653	0.786	4.650	<0.001***
sexM:age3:langM	−16.378	33.615	−0.487	0.626

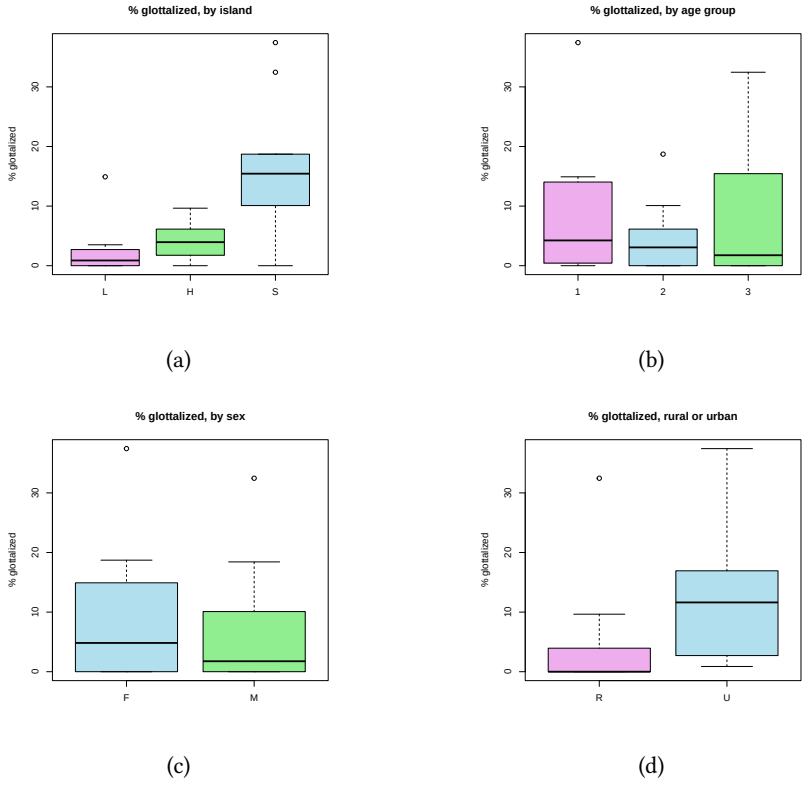


Figure 7: Glottalization rates by island (Lewis, Harris, Skye), age group (1 = 17–25, 2 = 35–45, 3 = 55+), sex, and rural vs. urban

While this lack of distinctive voicing has been noted to affect oral stops as well as sibilants in HE, Shuken's collected material was particularly well designed to illustrate the feature among sibilants, especially in word-final position. Her stimulus set includes 33 potential contexts for devoicing, including 27 instances of word-final /z/, 1 example of /z/ in initial position; 1 medial and 2 final examples of /dʒ/; and 1 example of medial /ʒ/ (Table 3). There were thus 2310 potential occurrences of sibilant devoicing in the analyzed data. Of these, 639 tokens actually exhibited devoicing, identified as the absence of visible pulses or a voice bar in the waveform and spectrogram (Figure 8).

Devoicing was somewhat more equitably distributed across the demographic groups than either preaspiration or glottalization (Table 8). That said, there are clear patterns all the same (Figure 9, Figure 10). First, devoicing seems to be more frequent among Lewis speakers than among those from Harris or Skye

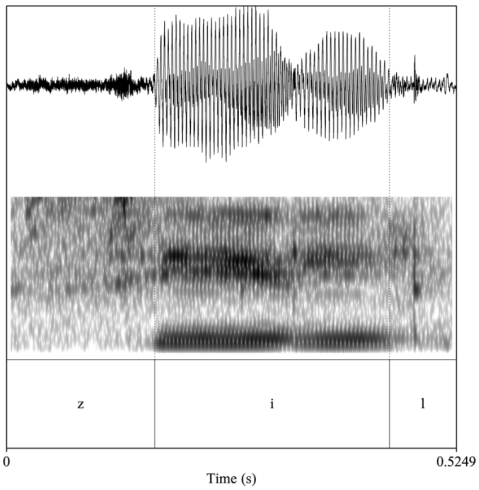


Figure 8: The word *zeal* as produced by H2, a rural bilingual female speaker from Harris in the 17–25 age group showing devoicing of the initial /z/.

(Figure 10a); more common in the young and middle age groups (Figure 10b); and more common among female than male speakers overall (Figure 10c).

Table 8: Distribution of devoiced tokens according to island, sex, age group.

	Lewis		Harris		Skye	
	M	F	M	F	M	F
Age 1 (17–25)						
Rural	39.4	56.1	36.4	22.7	na	na
Urban	29.3	44.7	na	na	15.2	38.4
Age 2 (35–45)						
Rural	45.5	48.5	15.2	21.2	na	18.2
Urban	24.2	53.5	na	na	4.5	39.4
Age 3 (55+)						
Rural	59.1	43.9	16.7	19.7	1.5	19.7
Urban	15.9	25.8	na	na	1.5	14.5

Statistical analysis of the devoicing data (Table 9) was conducted using a logistic mixed-effects model via the glmer function of the lme4 package. The final

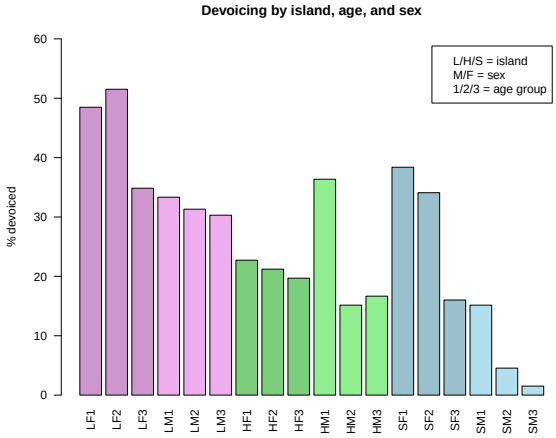


Figure 9: Rates of devoicing by according to island (Lewis, Harris, Skye), sex, and age group (1 = 17–25, 2 = 35–45, 3 = 55+)

model included *word* as random effect; and *island*, *age group*, *sex*, and *rural/urban* as fixed effects. Interactions between the fixed effects were also modeled. Overall, the model indicated that the effect of *island* was significant: the rate of devoicing was significantly higher in Lewis than in Harris (estimate $= -1.755$, $z = 4.195$, $p < 0.001$), though not in Skye (estimate $= -0.874$, $z = -1.360$, $p = 0.174$). We also see a general pattern of male speakers devoicing at significantly lower rates than female speakers in Lewis and in Harris, though not on Skye except among the older urban groups. The Skye result may be distorted by the absence of rural male Skye speakers in the younger and middle age groups, however.

5 Discussion

5.1 Summary of results

This analysis of a broad cross section of the recordings in the Shuken archive allows for a more detailed view of HE than was previously available. With respect to preaspiration and glottalization, this study both expands on Shuken (1984, 1985, 1986) and provides a baseline for comparison with studies using data collected several decades later (Clayton 2015, 2017, 2018). For example, in Clayton's later study, preaspiration could be found at least sporadically in most HE speakers, including those from islands not included in Shuken's sample. But of those speakers, only a small number of older rural women from Lewis provided more than a

Table 9: Mixed-effects model of complete devoicing results. Initial intercept represents younger female Lewis speakers.

Random effects		Variance	SD	
word		1.217	1.103	
Fixed effects	Estimate	SE	z	p
(Intercept)	0.247	0.337	0.733	0.464
islH	−1.755	0.418	−4.195	<0.001***
islS	−0.874	0.643	−1.360	0.174
age2	−0.376	0.388	−0.970	0.332
age3	−0.600	0.388	−1.544	0.123
rurbU	−0.562	0.337	−1.669	0.095
sexM	−0.826	0.390	−2.115	0.034*
islH:age2	0.276	0.592	0.466	0.641
islS:age2	−0.821	0.845	−0.972	0.331
islH:age3	0.394	0.596	0.661	0.509
islS:age3	−0.486	0.482	−1.010	0.312
islS:rurbU	0.560	0.570	0.981	0.327
age2:rurbU	0.812	0.488	1.664	0.096
age3:rurbU	−0.404	0.528	−0.765	0.444
islH:sexM	1.600	0.573	2.793	0.005**
islS:sexM	−2.485	1.650	−1.506	0.132
age2:sexM	0.677	0.548	1.235	0.217
age3:sexM	1.580	0.553	2.858	0.004**
rurbU:sexM	0.028	0.496	0.057	0.954
islS:age2:rurbU	0.436	0.852	0.512	0.609
islH:age2:sexM	−1.910	0.841	−2.271	0.023*
islS:age2:sexM	−0.724	0.732	−0.988	0.323
islH:age3:sexM	−2.583	0.842	−3.066	0.002**
islS:age3:sexM	−1.185	1.199	−0.988	0.323
islS:rurbU:sexM	1.845	1.585	1.164	0.244
age2:rurbU:sexM	−1.411	0.701	−2.012	0.044*
age3:rurbU:sexM	−1.467	0.743	−1.975	0.048*

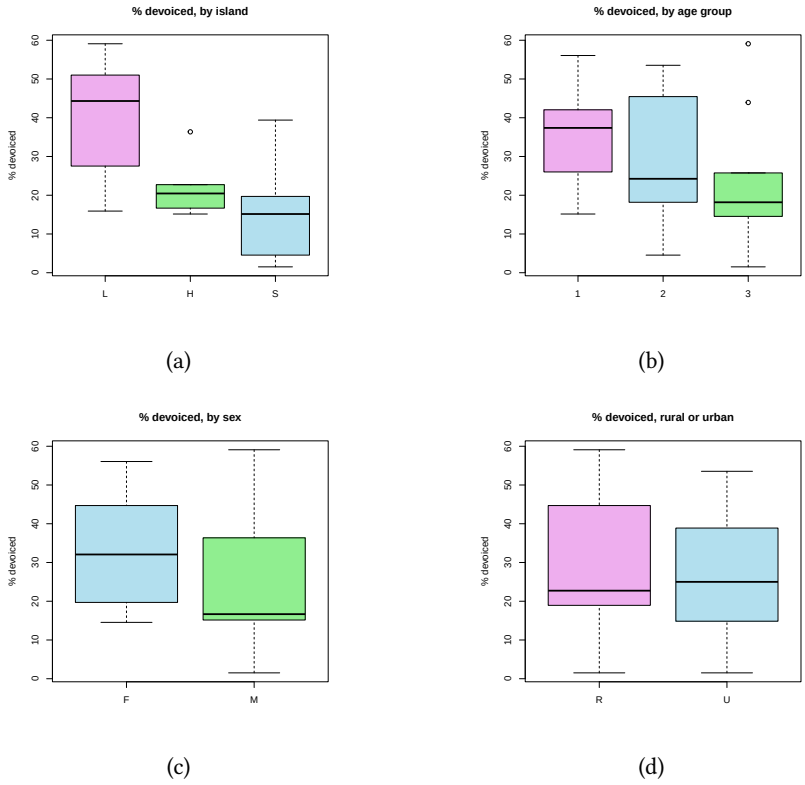


Figure 10: Devoicing rates by island, age group, sex, and rural vs. urban.

handful of preaspirated tokens. This could well be a development of the situation that Shuken (1985) discerned three decades previously, in which preaspiration predominated among participants from Lewis, particularly rural speakers, and particularly women. Though in that paper, Shuken did not consider the older two age groups from her full sample, the present analysis adds those two older age brackets, and finds that the same pattern holds: preaspiration is more abundant in Lewis than Harris or Skye, more abundant in rural areas compared to urban areas, and more abundant in women than men. The current study also finds that older speakers preaspirate more often than younger speakers. If we fast-forward three decades, so to speak, it is easy to see how the situation visible in Shuken's earlier data could readily develop into the state of affairs Clayton found in speakers three decades later.

Taken as a whole, the pattern we see with preaspiration is consistent with a conservative, local prestige feature, that is, one that speakers may exploit to

express affiliation with the local island community (cf. Labov 1963). Given the salience of preaspiration as a Gaelic feature, we may even suggest that the feature indexes a local Gaelic-speaking identity. However, with its evidently decreasing frequency, even among Lewis speakers, it is possible that preaspiration is becoming progressively less valuable to speakers in this regard.

On the other hand, the notably higher frequency of glottalization among Skye speakers, especially urban/monolingual speakers, instead bears the hallmarks of an incoming feature. Glottalization may well have been introduced, or reinforced, through contact with the mainland via newly arrived incomers from mainland urban communities like Glasgow or Aberdeen. Glottalization, particularly T-glottaling, is an expanding if stigmatized feature of working-class speech, in Scotland just as in England (Wells 1982, Trudgill 1988, Stuart-Smith 1999a,b). While we may expect a change favoring a non-standard form to be led by men (cf. Trudgill 1972, Labov 1963), especially one that confers working-class covert prestige, the lack of a significant difference by sex in the HE data is consistent with a similar lack of differentiation noted for T-glottaling in Glasgow English (Stuart-Smith 1999a), and indeed seems to presage a similar, later finding for HE in which glottalization shows no significant gender distinction (Clayton 2018).

Devoicing in certain respects patterns like preaspiration in the present study: it is more common among Lewis speakers than those from the other two islands (though not exclusive to Lewis), more common among female speakers than male speakers, and more abundant among rural speakers than urban speakers. On the other hand, devoicing is distinct from preaspiration in that it is noticeably more common among younger speakers than those in the middle or older age brackets (though statistically this age differentiation was only significant among men). The greater prevalence among women and among younger speakers is what we might associate with an incoming prestige feature (Labov 1990, Eckert 1989). But devoicing is not an incoming feature: it has been ascribed to the English of the Highland and Island region in one form or another for at least a century – sometimes represented as an anomaly that ought to be corrected (Grant 1913: 42), and sometimes caricaturized as in Compton Mackenzie's (1947) novel *Whisky Galore*. Indeed, one might expect there to be a stigma associated with the feature, as there is with T-glottaling; in fact, during their interviews, at least two female speakers among Shuken's participants singled out devoicing as an incorrect or problematic aspect of the local English. Thus, devoicing is not a prestige feature in the conventional sense. One wonders, though, whether devoicing is being exploited by younger speakers in lieu of preaspiration: as a local prestige feature indexing islander identity.

5.2 Further directions

There are two broad areas of further analysis and fieldwork suggested by the analyses presented here. First, the findings discussed in this chapter prompt many questions about the current status of HE. The body of material collected for the most recent published study of HE (Clayton 2015, 2017, 2018) illustrated only preaspiration and glottalization, and so can tell us little about the third variable considered in this chapter, sibilant devoicing. There is also the issue of geographic scope. While Clayton did consider material from a relatively broad array of islands within the Hebrides island group, apart from Skye and Lewis most of those islands were represented by only one or two speakers; no speakers were from the more southerly islands in the archipelago. Mull, Jura, Islay, Barra, and many other islands thus remain completely uninvestigated.

There is also a substantial quantity of analysis remaining to be carried out within the Shuken collection, some of which is currently in progress. First, the large and varied stimulus set allows the consideration of additional phonological features beyond those already considered, for example retroflex /rs/ and vowel epenthesis (indeed, analyses of these two features are currently in progress). Second, the reading passage material should be added. As a stylistically distinct form of language use (cf. Trudgill 1972), this material could well bring additional focus to those variables considered in the present chapter and any future analyses. Finally, only a limited set of social factors were considered in this chapter, including island of origin, rural vs urban settings, sex, age, and language ability. While these factors have allowed a number of insights into the distribution of the linguistic features considered, in particular their geographic distribution and relative abundance within the relevant demographic groups, as yet we know little about how preaspiration, glottalization, and devoicing may correlate with social factors such as education, economic class, or profession; or how speakers may exploit them to express group affiliation or construct identity.

Acknowledgements

I would like to thank the very helpful staff at the University of Edinburgh School of Scottish Studies archives, in particular Elliot Holmes and Kirsty Stewart, who made this work possible. I would also like to thank the two anonymous reviewers for their very helpful comments and suggestions on a previous draft of this chapter.

References

Aitken, Adam Jack. 1981. The Scottish vowel length rule. In Michael Benskin & M. L. Samuels (eds.), 131–157. The Middle English Dialect Project.

Baayen, R. Harald, Douglas Davidson & Douglas Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). 390–412. DOI: 10.1016/j.jml.2007.12.005.

Barnes, Michael. 1984. Orkney and Shetland norn. In Peter Trudgill (ed.), *Language in the British Isles*, 352–366. Cambridge: Cambridge University Press.

Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67. 1–48. DOI: 10.18637/jss.v067.i01.

Bird, Barbara. 1997. Past and present studies of Hebridean English phonology. In Hildegard L.C. Tristram (ed.), *The Celtic Englishes* (Anglistische Forschungen 247), 287–300. Heidelberg: Carl Winter.

Boersma, Paul & David Weenink. 2020. *Praat: doing phonetics by computer (version 6.1.05)*. Computer program. Retrieved from http://www.praat.org/.

Borgstrøm, Carl H.J. 1940. *The dialects of the Outer Hebrides: A linguistic survey of the Gaelic dialects of Scotland*, vol. I (Norsk Tidsskrift for Sprogvidenskap). Oslo: H. Aschehoug & Co.

Bosch, Anna & Kenneth de Jong. 1997. The prosody of Barra Gaelic epenthetic vowels. *Studies in the Linguistic Sciences* 27(1). 1–15.

Catford, John C. 1957a. Shetland dialect. *Shetland Folk Book* 3. 71–75.

Catford, John C. 1957b. The linguistic survey of Scotland. *Orbis* 6. 105–121.

Clayton, Ian. 2010. *On the natural history of preaspirated stops*. University of North Carolina, Chapel Hill. (Doctoral dissertation).

Clayton, Ian. 2015. Preaspirated stops in the English of Scottish Gaelic-English bilinguals. In The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of the Phonetic Sciences*. Glasgow, UK: University of Glasgow.

Clayton, Ian. 2017. Preaspiration in Hebrides English. *Journal of the International Phonetic Association* 47(2). 155–181. DOI: 10.1017/S002510031700007X.

Clayton, Ian. 2018. Revisiting Hebrides English. *English World-Wide* 19(2). 157–189. DOI: 10.1075/eww.00009.cla.

Clement, R.D. 1980. Highland English. *Scottish Literary Journal* Supplement 12. 13–18.

Clement, R.D. 1997. Highland English. In Hildegard L.C. Tristram (ed.), *The Celtic Englishes* (Anglistische Forschungen 247). Heidelberg: Carl Winter.

Dorian, Nancy. 1978. *East Sutherland Gaelic: The dialect of the Brora, Golspie, and Embo fishing communities*. Dublin: University of Pennsylvania Press.

Eckert, Penelope. 1989. The whole woman: Sex and gender differences in variation. *Language Variation and Change* 1. 245–267. DOI: 10.1017/s095439450000017x.

Gillies, William. 2009. Scottish Gaelic. In Martin J. Ball & Nicole Müller (eds.), *The Celtic languages*, 230–304. London, New York: Routledge. DOI: 10.4324/9780203882481.

Grant, William. 1913. *The pronunciation of English in Scotland*. Cambridge: Cambridge University Press.

Hammond, Michael, Natasha Warner, Andréa Davis, Andrew Carnie, Diana Archangeli & Muriel Fisher. 2014. Vowel insertion in Scottish Gaelic. *Phonology* 31. 123–153. DOI: 10.1017/s0952675714000050.

Hickey, Raymond. 2013. Supraregionalisation and dissociation. In J. K. Chambers & Natalie Schilling (eds.), *Handbook of language variation and change*, 537–554. Wiley-Blackwell. DOI: 10.1002/9781118335598.ch25.

Highlands & Islands Enterprise. 2007. *Evaluation of the economic and social impacts of the Skye bridge*. https://hitrans.org.uk/Documents/Evaluation_of_the_Economic_and_Social_Impacts_of_the_Skye_Bridge.pdf/m=1 (5 June, 2025).

Highlands & Islands Enterprise. 2020. *Highlands and Islands Area Profiles 2020: Na h-Eileanan an Iar [The Western Isles]*. https://www.hie.co.uk/media/10594/na-h-eileanan-an-iar-area-profile-2020.pdf (5 June, 2025).

Holmer, Nils. 1957. *The gaelic of Arran*. Dublin: Dublin Institute for Advanced Studies.

Holmer, Nils. 1981. *The gaelic of Kintyre*. Dublin: Dublin Institute for Advanced Studies.

Iosad, Pavel, Michael Ramsammy & Patrick Honeybone. 2015. *Preaspiration in North Argyll Gaelic and its contribution to prosodic structure*. Forum for Research on the Languages of Scotland and Ulster, Conference date: 12-08-2015 Through 14-08-2015.

Johnston, Paul. 1997. Regional variation. In Charles Jones (ed.), *The Edinburgh history of the Scots language*, 433–513. Edinburgh University Press. DOI: 10.1515/9781474410977-013.

Jones, Charles. 2002. *The English language in Scotland: An introduction to Scots*. East Linton, Scotland: Tuckwell Press.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. Lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). 1–26. DOI: 10.32614/cran.package.lmertest.

Labov, William. 1963. The social motivation of a sound change. *Word* 19. 273–309. DOI: 10.1080/00437956.1963.11659799.

Labov, William. 1990. The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2. 205–254. DOI: 10.1017/s0954394500000338.

Ladefoged, Peter, Jenny Ladefoged, Alice Turk, Kevin Hind & St. John Skilton. 1998. Phonetic structures of Scottish Gaelic. *Journal of the International Phonetic Association* 28. 1–41. DOI: 10.1017/s0025100300006228.

Mackenzie, Compton. 1947. *Whisky galore*. London: Chatto & Windus.

Mather, James & H. H. Speitel. 1986. *The linguistic atlas of Scotland: Scots section.* Vol. 3: Phonology. Hamden, Connecticut: Croom Helm Ltd.

McClure, J. Derrick. 1994. English in Scotland. In Robert Burchfield (ed.), *The Cambridge history of the English language, English in Britain and overseas: Origins and development*, 23–93. Cambridge: Cambridge University Press. DOI: 10.1017/CHOL9780521264785.003.

Melchers, Gunnel. 2008. English spoken in Orkney and Shetland: Morphology, syntax, and lexicon. In Bernd Kortmann & Clive Upton (eds.), *Varieties of English 1: The British Isles*, 285–298. Mouton de Gruyter. DOI: 10.1515/9783110208399.2.285.

Nance, Claire. 2014. Phonetic variation in Scottish Gaelic laterals. *Journal of Phonetics* 47. 1–17. DOI: 10.1016/j.wocn.2014.07.005.

Nance, Claire & Jane Stuart-Smith. 2013. Pre-aspiration and post-aspiration in Scottish Gaelic stop consonants. *Journal of the International Phonetic Association* 43. 129–152. DOI: 10.1017/s0025100313000042.

National Records of Scotland. 2023. *Mid-year Population Estimates*. https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/mid-year-population-estimates/population-estimates-time-series-data (5 June, 2025).

Ní Chasaide, Ailbhe. 1985. *Preaspiration in phonological stop contrasts*. University College of North Wales, Bangor. (Doctoral dissertation).

Ó Dochartaigh, Cathair. 1994-1997. *Survey of the Gaelic dialects of Scotland*, vol. 5 vols. Edinburgh: Dublin Institute for Advanced Studies.

Ó Giollagáin, Conchúr, Gòrdan Camshron, Pàdruig Moireach, Iain Caimbeul, Brian MacDonald & Tamás Péterváry. 2020. *The Gaelic crisis in the vernacular community: A comprehensive sociolinguistic survey of Scottish Gaelic*. Aberdeen: Aberdeen University Press. DOI: 10.57132/mpub.14497417.

Ó Murchú, Máirtín. 1989. *East Perthshire Gaelic*. Dublin: Dublin Institute for Advanced Studies.

Oftedal, Magne. 1956. *The Gaelic of Leurbost, Isle of Lewis: A linguistic survey of the Gaelic dialects of Scotland*, vol. 3 (Norsk Tidskrift for Sprogvidenskap). Oslo: Aschehoug & Co.

Sabban, Annette. 1985. On the variability of Hebridean English syntax: The verbal group. In Manfred Görlach (ed.), *Varieties of English around the world: Focus on Scotland*, 125–143. Amsterdam: John Benjamins. DOI: 10.1075/veaw.g5.09sab.

Shuken, Cynthia. 1979. Aspiration in Scottish Gaelic stop consonants. In Harry Hollien & Patricia Hollien (eds.), *Current issues in the phonetic sciences: Proceedings of the IPS-77 congress, Miami Beach, Florida, 17–19 December 1977*, 451–458. Amsterdam: Benjamins. DOI: 10.1075/cilt.9.48shu.

Shuken, Cynthia. 1984. Highland and Island English. In Peter Trudgill (ed.), *Language in the british isles*, 152–166. Cambridge: Cambridge University Press.

Shuken, Cynthia. 1985. Variation in Hebridean English. In Manfred Görlach (ed.), *Varieties of English around the world*, 145–158. Amsterdam: John Benjamins. DOI: 10.1075/veaw.g5.

Shuken, Cynthia. 1986. Vowel systems in Hebridean English. In Donald MacAulay & J. Derrick McClure (eds.), *Proceedings of the first international conference on the languages of Scotland*, 131–139.

Smith, Norval. 1999. A preliminary account of some aspects of Leurbost Gaelic syllable structure. In Harry van der Hulst & Nancy A. Ritter (eds.), *The syllable: Views and facts*, 577–630. Mouton de Gruyter. DOI: 10.1515/9783110806793.577.

Speitel, H. H. 1981. The geographic position of the Scots dialect in relation to the Highlands of Scotland. In Michael Benskin & M.L. Samuels (eds.), *So meny people longages and tonges: Philological essays in Scots and mediaeval English presented to Angus McIntosh*, 131–157. Middle English Dialect Project.

Stuart-Smith, Jane. 1999a. Glasgow: Accent and voice quality. In Paul Foulkes & Gerard Docherty (eds.), *Urban voices: Accent studies in the British Isles*, 203–222. Routledge. DOI: 10.4324/9781315832470-19.

Stuart-Smith, Jane. 1999b. Glottals past and present: A study of T-glottalling in Glaswegian. *Leeds Studies in English* 30. 181–204.

Stuart-Smith, Jane. 2003. The phonology of modern urban Scots. In John Corbett, J. Derrick McClure & Jane Stuart-Smith (eds.), *The Edinburgh companion to Scots*, 110–137. Edinburgh: Edinburgh University Press. DOI: 10.1515/9781474421591-010.

Stuart-Smith, Jane. 2008. Scottish English: Phonology. In Bernd Kortmann & Clive Upton (eds.), *Varieties of English: The British Isles*, 48–70. Mouton de Gruyter. DOI: 10.1515/9783110208399.1.48.

Ternes, Elmar. 1973. *The phonemic analysis of Scottish Gaelic, based on the dialect of Applecross, Ross-shire*. Hamburg: Helmut Buske Verlag.

Trudgill, Peter. 1972. Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society* 1. 179–195. DOI: 10 . 1017 / s0047404500000488.

Trudgill, Peter. 1988. Norwich revisited: Recent linguistic changes in an English urban dialect. *English World-Wide* 9(1). 33–49. DOI: 10.1075/eww.9.1.03tru.

Wells, John C. 1982. *Accents of English 2: The British Isles*. Cambridge: Cambridge University Press.

Withers, Charles. 1984. *Gaelic in Scotland 1698-1981: The geographic history of a language*. Edinburgh: John Donald.

Chapter 8

Gestural timing and contrast: An Irish case study

Ryan Bennett[a], Jaye Padgett[a], Grant McGuire[a], Máire Ní Chiosáin[b] & Jennifer Bellik[a]

[a]University of California, Santa Cruz [b]University College Dublin

Though a palatalized vs. velarized (*slender* vs. *broad*) contrast is fundamental to Goidelic languages of the Celtic family, contrasts involving either palatalization or velarization are typologically uncommon. Across languages, such contrasts are often dispreferred in onset compared to coda position, and in labial compared to coronal consonants. Previous phonetic research on both Russian and Irish has suggested that these asymmetries in contrast loss might be explained by, or at least correlated with, asymmetries in articulatory reduction and timing. We explore the timing of palatalization and velarization gestures in Irish stop consonants, and find differences between onset vs. coda position consistent with this general claim. However, we also find that gestural timing is structured in ways that allow for perceptual cues to the contrast to be recoverable, including in ways that might facilitate perception of the contrast in codas and (possibly) in labial consonants. Put simply, gestural timing in Irish may be optimized to support vulnerable contrasts.

1 Introduction

The living Celtic languages are divided into two groups, a Goidelic branch (including Irish and Scottish Gaelic) and a Brittonic branch (including Welsh and Breton).[1] In terms of phonological properties, an important feature distinguishing Goidelic languages from Brittonic ones is a phonemic contrast between palatalized and non-palatalized consonants, traditionally called *slender* and *broad* respectively. For example, compare the Irish minimal pair *brád* /bʸrʸɔːdʸ/ 'drizzle'

[1]For a recent discussion of the classification of Celtic languages, see Eska (2017).

Ryan Bennett, Jaye Padgett, Grant McGuire, Máire Ní Chiosáin & Jennifer Bellik. 2025. Gestural timing and contrast: An Irish case study. In Andrew Carnie, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond & Luis Irizarry (eds.), *Foundational approaches to Celtic linguistics*, 283–341. Berlin: Language Science Press. DOI: 10.5281/zenodo.15654871

and *bráid* /bˠrˠɔːdʲ/ 'neck, throat'; the two words differ only in whether the final consonant is palatalized or velarized. (These terms are explained below.) This palatalization contrast developed in Proto/Old Irish (Greene 1973; McCone 1996; Stifter 2017), and it occurs also in Scottish Gaelic (Ternes 2006; Nance & Ó Maolalaigh 2021) and Manx (Lewin 2019).

Many phonemic contrasts are allowed only in certain phonological environments. For example, though English distinguishes voiced obstruents from voiceless ones both word/syllable-initially and -finally (e.g., *back* vs. *pack* and *cab* vs. *cap*, Russian prohibits the contrast finally, neutralizing in favor of the voiceless series: compare [ˈlˠugˠə] 'meadow (GEN.)' and [ˈlˠukˠə] 'onion (GEN.)' (from /lˠugˠ-a/ and /lˠukˠ-a/, respectively) to [lˠukˠ] 'meadow' and [lˠukˠ] 'onion' (from /lˠugˠ/ and /lˠukˠ/, respectively).

It has long been understood that phonological alternations, including those leading to the positional neutralization of contrast, might have historical origins in phonetic processes, a hypothesis pursued most notably by Ohala (e.g. Ohala 1981; 1993b), but since explored by many. One way to test a hypothesis about the phonetic origins of positional contrast and neutralization is to examine a language that does *not* neutralize the relevant contrast, and look for potential phonetic seeds of neutralization there. Indeed, we have done this for Irish (see discussion further on). However, there is also much work showing that languages can be structured in ways that militate *against* contrast loss, a view most notably pursued by Lindblom (1986; 1990), but again explored since then by many.

In this paper, we explore details of the timing of palatalization and velarization gestures in Irish stop consonants as a function of prosodic position (word/syllable-initial vs. -final) and place of articulation (labial, coronal, dorsal). We find that timing patterns vary more in word/syllable-final position, a fact that may help explain the vulnerability of this contrast to loss in that position. However, we also find evidence that timing in that position can be structured in ways that might *facilitate* perception of the palatalization contrast. As for place of articulation, we find that patterns of timing in labial consonants, which are also vulnerable to loss of a palatalization contrast, are more consistent than for other places of articulation. This may likewise facilitate perception of the palatalization contrast, though here the facts are less conclusive. Overall, our results suggest that gestural timing in Irish may be optimized to support contrasts that are vulnerable.

2 Irish and the typology of contrastive palatalization

2.1 Palatalization and velarization

Palatalized (slender consonants [C^{j}] are produced with a secondary high-front tongue constriction, akin to the constriction for a palatal vowel [i] or glide [j], which is roughly simultaneous with the primary constriction for that consonant.[2] Velarized (broad) consonants [C^{ɣ}] are produced with a secondary tongue body constriction that is high and back in the mouth, as for [ɯ u] or [ɥ w], or sometimes in the mid-back or uvular region, as for [ɤ o] (International Phonetic Association 1999; Kochetov 2002; Bennett et al. 2018; Shaw et al. 2021). In this article, the term 'secondary articulation' encompasses palatalization and velarization as defined above.

2.2 The palatalization contrast in Irish

All consonants in Irish are contrastively palatalized or velarized. Table 1 illustrates these phonemic contrasts with the basic inventory of consonants in Connacht Irish (or more specifically Connemara Irish), which here represents the western dialect group (De Bhaldraithe 1975; Ó Siadhail 1991; Ní Chiosáin 1991; 1994; 1999; Ní Chiosáin & Padgett 2012). The specific phonetic form of palatalized and velarized consonants varies somewhat across dialects (e.g. Ní Chasaide 1995; Hickey 2011), but all varieties of Irish share fundamentally the same inventory of phonological contrasts illustrated in Table 1.

As illustrated in Table 1, /C^{ɣ} C^{j}/ contrasts are maintained across all places of articulation in Irish. In addition, /C^{j}/ and /C^{ɣ}/ can appear across all vowel contexts as well as word/syllable-initially and -finally (some examples given in Table 2, left). In the data analyzed in this paper, word-initial sounds are also syllable-initial (and vice-versa), and word-final sounds are also syllable-final (and vice versa). For ease of reference we will often use 'onset' to refer to the former and 'coda' to refer to the latter.[3]

[2]Since not all languages with a palatalization contrast are reported to oppose velarization to palatalization, we use '/C$^{(ɣ)}$ C^{j}/' to refer to palatalization contrasts generally. We use capital 'C' to refer to any consonant, and capital 'P', 'T', 'K' to refer generically to labial, coronal, and dorsal consonants, respectively. Slash brackets '/ /' indicate abstract phonemes, while square brackets '[ ]' indicate surface phonetic forms. Transcriptions are broad: for example, coronal [d^{ɣ} d^{j}] are phonetically closer to dentalized [d̪^{ɣ}] and retracted [d̠^{j}] in e.g. Gaoth Dobhair Irish (Ní Chasaide 1995).

[3]Consonant clusters typically agree in their secondary articulations in Irish (Ní Chiosáin 1991; 1999). Such clusters may be palatalized (e.g. *coirpeach* [k^{ɣ}ir^{j}p^{j}əx^{ɣ}] 'criminal') or velarized (e.g. *léargas* [l^{j}e:r^{ɣ}g^{ɣ}əs] 'insight'). So, although word-internal codas may not have *independent* /C^{ɣ} C^{j}/

Table 1: Phonemic consonant inventory of Connemara Irish. Sounds in parentheses have highly restricted distributions.

	Labial	Coronal	Dorsal	Glottal
Stop	$p^ˠ$ $p^ʲ$	$t^ˠ$ $t^ʲ$	$k^ˠ$ $k^ʲ$	
	$b^ˠ$ $b^ʲ$	$d^ˠ$ $d^ʲ$	$g^ˠ$ $g^ʲ$	
Fricative	$f^ˠ$ $f^ʲ$	$s^ˠ$ $s^ʲ$	$x^ˠ$ $x^ʲ$	$h^ˠ$ ($h^ʲ$)
	$v^ˠ$ $v^ʲ$		($ɣ^ˠ$) ($ɣ^ʲ$)	
Nasal	$m^ˠ$ $m^ʲ$	$n^ˠ$ $n^ʲ$	$ŋ^ˠ$ $ŋ^ʲ$	
Liquid		$l^ˠ$ $l^ʲ$		
		$r^ˠ$ $r^ʲ$		

Table 2: Words showing distribution of /$C^ˠ$ $C^ʲ$/ (left) and minimal pairs (right).

caoin	[$k^ˠiːn^ʲ$]	'lament'	*pán*	[$p^ˠɔːn^ˠ$]	'pawnshop'
tiús	[$t^ʲuːs^ˠ$]	'thickness'	*peann*	[$p^ʲɔːn^ˠ$]	'pen'
láib	[$l^ˠɔːb^ʲ$]	'mud, muck'	*cat*	[$k^ˠat^ˠ$]	'cat (SG.)'
píob	[$p^ʲiːb^ˠ$]	'pipe (GEN.PL.)'	*cait*	[$k^ˠat^ʲ$]	'cat (PL.)'

The wide distribution of the /$C^ˠ$ $C^ʲ$/ contrast across all of these contexts allows for many minimal pairs (Table 2, right). It is also what makes the Irish language a useful and important testing ground for phonetically-oriented explanations of recurrent distributional restrictions on /$C^{(ˠ)}$ $C^ʲ$/ contrasts across languages.

2.3 Typological asymmetries in palatalization contrasts

Though allophonic secondary palatalization is found in many languages (Bateman 2011), its *contrastive* use, as in Irish /$b^ˠr^ˠɔːd^ˠ$/ vs. /$b^ˠr^ˠɔːd^ʲ$/, is relatively uncommon. Furthermore, among languages that do distinguish /$C^{(ˠ)}$ $C^ʲ$/, the contrast is not equally likely in all phonological environments (Takatori 1997; Kochetov

specifications, given that secondary articulations must assimilate in clusters, word-internal codas may be either palatalized or velarized just like word-final codas. (Compounds and prefixes allow disagreeing clusters, e.g. *inphosta* [$in^ʲ$-$f^ˠoːs^ˠt^ˠə$] 'marriable', *fadtéarma* [$f^ˠad^ˠ$-$t^ʲeːr^ˠm^ˠə$] 'longterm'; Ní Chiosáin 1991: Ch.2.)

The backness of short vowels is predictably determined by the secondary articulations of neighboring consonants, which is why we focus on long vowels in our discussion of vocalic contexts. See Section 4.3, Ní Chiosáin (1991), and Ó Siadhail (1991) for details.

2002; Iskarous & Kavitskaya 2018). First, such contrasts tend to be neutralized or lost diachronically in codas, as in Bulgarian (e.g. Scatton 1993; Iskarous & Kavitskaya 2018) or Nenets (e.g. Kochetov 2002). As a result, the presence of a /C$^{(ɣ)}$ C^{j}/contrast in coda position normally entails a corresponding contrast in onset position (but not vice-versa).[4] Second, /C$^{(ɣ)}$ C^{j}/contrasts tend to be neutralized or lost diachronically in labial consonants compared to coronals. For example, this pattern occurs in Czech (Short 1993; Iskarous & Kavitskaya 2018). These two restrictions may also interact, as in Belarusian, where palatalization is contrastive for both labials and coronals in syllable onsets, but palatalized labials do not occur in syllable codas (e.g. Bird & Litvin 2021).

Some signs of these asymmetries are evident within Celtic. Scottish Gaelic may have lost, or be in the process of losing, its palatalization contrast among labial consonants, with remnants in the form of /Pj/ (labial consonant – glide) sequences (Jackson 1967; De Búrca 1977; Ternes 2006; cf. Nance & Ó Maolalaigh 2021), as is true in some Polish dialects (Czaplicki 2010; Iskarous & Kavitskaya 2018). The same may have been true of Classical Manx (Lewin 2019). In Irish, there are many fewer words ending in palatalized labial stops compared to palatalized coronal stops, and the proportion of word-final labial stops that are palatalized is low compared to the proportion of word-final coronal stops that are palatalized, based on either a dictionary or corpus search of word types (Ní Chiosáin & Padgett 2012). As Kochetov (2002) notes, discussing similar facts for Russian, such asymmetries in frequency may reflect lexical attrition of contrasts in more vulnerable positions. They might, in turn, also enable further attrition of contrasts and possibly lead to neutralization, because language learners will have to learn the contrast based on fewer relevant lexical items.

In spite of such asymmetries in segment frequency, Irish still maintains the palatalization contrast in codas and in labials. For this reason, Irish is an ideal testing ground for hypotheses about the phonetic bases of contrast and contrast neutralization. For example, if contrasts are lost in certain contexts due to systematic phonetic asymmetries, then we might expect to find such phonetic asymmetries in Irish, even though it maintains the contrast. The hunt for phonetic precursors to neutralization of a /C$^{(ɣ)}$ C^{j}/ contrast has precedents, including work of our own (Kochetov 2002; 2005; 2006; Ní Chiosáin & Padgett 2012; Stoll 2017; Padgett & Ní Chiosáin 2018; Iskarous & Kavitskaya 2018; Kirkham & Nance 2022; Padgett et al.

[4]Marginal exceptions to this typological generalization include Estonian and Romanian, in which phonological palatalization only occurs word-finally or before a consonant (i.e. only in coda position). These unusual distributions owe to contextual palatalization before /i/, rendered opaque by historical or synchronic vowel deletion (Operstein 2010; Chitoran 2013; Malmi et al. 2023).

2023; Bennett et al. 2024). On the other hand, if languages are structured in ways that *support* contrast, then we might expect to find ways in which this is true for the Irish palatalization contrast. The work presented here explores both of these possibilities, focusing on the timing of palatalization and velarization with respect to a primary place constriction. It develops ideas first presented by Padgett et al. (2023). Here we bring new analyses to bear, including analyses of the gestural trajectories of individual tokens and an exploration of individual subject behavior. We develop the idea of perceptual optimization, and we consider and reject an alternative understanding of gestural timing that relies on articulatory principles along the lines of Sproat & Fujimura (1993) and Krakow (1999).

3 Gestural timing and contrast

3.1 Gestural organization and perception

Perceptual studies on Irish and Russian have found that listeners have more difficulty identifying or distinguishing /C$^{\gamma}$ C^{j}/ in coda position compared to onset position, and for Russian, in labials compared to coronals. Kochetov (2002; 2004) found that Russian listeners misidentified /p$^{\gamma}$ p^{j} t$^{\gamma}$ t^{j}/ more often in word/syllable-final position than in initial position, even though the relevant consonant was intervocalic due to a following vowel-initial word.[5] In addition, listeners misidentified /p^{j}/ more often than the other three consonants in coda position, most often confusing it with /p$^{\gamma}$/. Ní Chiosáin & Padgett (2012), Padgett & Ní Chiosáin (2018) found that Irish listeners discriminate /C$^{\gamma}$ C^{j}/ more poorly in coda position compared to onset position, for a number of stop and fricative consonants. (The codas in these stimuli were phrase-final.) As can be seen, these perceptual asymmetries mirror the typological ones discussed earlier, raising the possibility that the perceptual asymmetries *explain* the typology. In contexts where listeners cannot perceive the /C$^{\gamma}$ C^{j}/ contrast well, we might expect attrition and eventual loss of that contrast. We might also expect such perceptual asymmetries to have some basis in productions that are weaker, more variable, etc.

However, it's possible that languages are structured to be perceptually optimizing, in which case we might have a countervailing expectation: production of a /C$^{\gamma}$ C^{j}/ contrast might vary by position (initial or final) or by place (labial or coronal) in ways that *support* the contrast, specifically in vulnerable contexts. This idea and the previous are not contradictory; rather, there is strong reason

[5]There is no evidence that consonants resyllabify across words in Russian. For example, word-final consonants devoice even when vowel-initial words follow (see discussion in Padgett & Myers 2014). See also the discussion in Section 4.3.

to believe both. Which factor prevails in a given case, leading to loss or preservation of a contrast, depends on many factors that are generally beyond our power to predict.

The idea that languages might be structured in ways that facilitate the maintenance of phonological contrast goes back decades (e.g. Martinet 1952; 1964; Liljencrants & Lindblom 1972; Lindblom 1986; 1990). Most relevant here is work more specifically suggesting that *gestural timing* can be organized in ways that are perceptually optimizing. An important precedent is Kingston (1990), who argues that across languages glottal articulations are 'bound' in time most often to the release of stop consonants, because the burst characteristic of stops renders the glottal gesture most perceptually effective. Sonorant and fricative consonants lack bursts and so do not coordinate glottal articulations in the same way. Silverman (1997) argues in a similar way that laryngeal states creating phonation and tone contrasts in vowels are timed in nuanced ways that facilitate their perceptibility, and Wright (1996) argues that place gestures in consonant clusters are timed so as to ensure recoverability of burst cues in consonants that lack internal place cues or formant transitions. More generally, Kingston & Diehl (1994) argue that speakers systematically control articulations in ways that support the perceptual distinctiveness of contrast, and there is a great deal of other evidence that phonetic targets are ultimately auditory, not gestural. (For an overview and arguments, see Kingston 2019.)

The above observations do not necessarily imply that speakers *intentionally* structure their articulations to make their speech clear or intelligible for listeners. While that *could* be the case, it is also plausible that articulatory patterns which are easier to hear are also easier to learn, and therefore more likely to persist over time (e.g. Ohala 1993a). Articulatory patterns which produce less robust acoustic outcomes are more likely to be misheard by listeners: gradually, such misperceptions can lead to sound changes involving the collapse of contrasts (e.g. Ohala 1981 and many others). In this way, perceptually-supportive articulatory strategies can emerge organically, without any intent on the part of speakers. That said, we remain open to the possibility that speakers do in fact strive to produce clear, intelligible speech in at least some circumstances. In either case, articulatory timing clearly plays an important role in the perceptibility of contrasts.

3.2 Gestural timing in palatalization contrasts

Previous work on the timing of secondary articulations in both Irish and Russian suggests that the tongue body gesture associated with palatalization reaches its

target (or peaks) around the release of the primary place constriction for onset consonants (e.g. Ladefoged & Maddieson 1996: Ch.10; Kochetov 2002; 2006; Iskarous & Kavitskaya 2010; Bennett et al. 2018; Padgett et al. 2023). More specifically for Irish, Bennett et al. (2018) find that tongue body fronting for palatalized /C^{j}/ tends to peak at C release in both stops and fricatives. In contrast, velarized /C^{ɣ}/ does not show any clear, consistent asymmetries across time points: the magnitude of dorsal backing tends to be comparable at the beginning of C constriction, midpoint, and release.

Bennett et al. (2018) only examine consonants that are onsets (and prevocalic), and so do not address the timing of secondary dorsal constrictions in codas (or postvocalic position). However, based on an electromagnetic midsagittal articulometry (EMMA) study of four speakers, Kochetov (2006) found that Russian /p^{j}/ was produced differently in onset position compared to coda position. Kochetov measured the lag between the achievement of the secondary and primary articulations (defined in terms of velocity minima) of /p^{j}/, as well as the lag between the release of the two gestures. For all speakers, in onset position both the achievement and release of the palatalization gesture followed that of the labial constriction by 5–55 ms on average (depending on the speaker). This is consistent with our findings for Irish mentioned above. In coda position, the relative timing of the palatalization gesture with respect to the labial gesture was found to be more variable by speaker, but it generally occurred earlier than in onset position. For example, the achievement of the palatalization gesture preceded that of the labial gesture by 5–25 ms for three of the speakers, but followed it for the fourth speaker. Biteeva (2021: 159–160) reports very similar findings for /t^{j}/, based on another EMMA study of nine speakers. Both studies characterize the relationship between primary and secondary articulations as sequential for onsets and more simultaneous for codas.

3.3 Gestural timing and perceptual consequences

It is worth considering the timing facts above in light of their effects on the production of cues to palatalization. As Kochetov (2006) points out, it is perceptually adaptive for the palatalization gesture of a stop to lag its primary place gesture in onset consonants. There are two major cues to the palatalization status of a stop: one is in the formant transitions between the consonant and a neighboring vowel (primarily the second formant), and the other is in the quality of the consonant burst, the spectral properties of which are shaped by the presence of palatalization or velarization. In onset consonants, these distinct cues more or less coincide in time at the release of the primary place gesture, indicated by

the leftmost oval in Figure 1. Following the reasoning of Kingston (1990) for the timing of glottal articulations with respect to stop consonants (discussed earlier), coordinating palatalization to peak around this location maximizes its contrast potential. This fits well with the general finding that onset consonant-vowel transitions are perceptually robust (Fujimura et al. 1978; Ohala 1990; Steriade 2001; Wright 2004; Kochetov 2006).

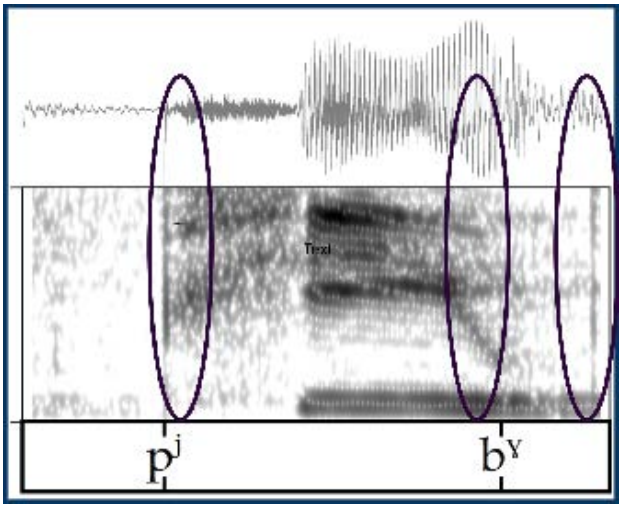


Figure 1: Waveform and spectrogram of the word *píob* /pʲiːbˠ/ (Connacht Speaker 1 repetition 1), showing location of major cues to Irish secondary articulations in onset (left oval) and coda (right two ovals) stops.

Things are different for coda stops, where the two major cues to palatalization are split in time, as shown on the right of Figure 1. The formant transition cues to the contrast precede the stop closure (note the salient drop in the second formant for /bˠ/), while the burst cues lie at the release of the stop constriction (rightmost oval in the figure). If secondary articulations are timed to be perceptually optimizing, what should that entail for stops in final position? We do not know enough about the perception of palatalization cues to answer this question definitively. However, a reasonable first thought would be that gestures should be coordinated in a way that facilitates the perception of *both* formant transitions and burst properties. This would be true if the secondary articulation were to be achieved at, or slightly before, the onset of the primary articulation and be released at, or slightly after, the release of the primary articulation.[6]

[6]The discussion here is oversimplified, since it ignores the larger phrasal context of the word in Figure 1. We return to this issue in the paper's discussion and conclusion.

As discussed in the last section, previous work on Russian has suggested that the primary and secondary articulations of coda palatalized stops can be roughly simultaneous, consistent with the above prediction, or that the palatalization gesture slightly precedes the primary one, perhaps favoring the formant transition cues. Roughly simultaneous coordination has also been found for Russian /lʲ/ (Kochetov 2005) and /sʲ/ (Kochetov 2009). In these studies, the sounds occurred word/syllable-finally before a vowel-initial word. As Kochetov (2005) notes for /lʲ/, these facts also seem consistent with the hypothesis about timing and perception: since these sounds have internal cues to palatalization (that is, the palatalized state of the consonant is audible during the steady state portion of its production), have no bursts, and would have formant transitions on both sides, we have no reason to expect timing of the palatalization gesture to be coordinated with either the beginning or end of the consonant in particular, regardless of syllable position (see also Gick et al. 2006).[7]

However, in the same phonetic context (coda before a vowel-initial word), Kochetov (2009) found that the palatalization gesture of /tʲ/ was timed to coincide more with the *release* of the primary coronal gesture. Assuming such a difference turned out to be systematic, it might suggest that gestural timing is sensitive not just to the presence, but to the relative salience, of phonetic cues. Ní Chiosáin & Padgett (2012) report that Irish /tʲ dʲ/ differ more from /tˠ dˠ/, respectively, in burst duration, intensity and center of gravity than /pʲ bʲ/ differ from /pˠ bˠ/. Indeed, in both Irish and Russian the release of /tʲ dʲ/ is very salient, even sounding somewhat affricated. In contrast, /pʲ bʲ/ differ *more* from /pˠ bˠ/, respectively, in second formant values than /tʲ dʲ/ differ from /tˠ dˠ/, in both Irish and Russian (Purcell 1979; Ní Chiosáin & Padgett 2012). These considerations might suggest that, if the timing of secondary articulation gestures in coda consonants is to favor either the beginning or end of the primary gesture, it would be more adaptive to favor the beginning in the case of labial stops and the end in the case of coronal stops. However, it should be noted that Kochetov (2009) found coordination with the end also in the case of Russian coda /nʲ/. This obviously does not sit well with the above reasoning, and overall this remains fairly speculative given the amount of available data.[8]

[7] In this paper we may use the terms 'start' and 'beginning' interchangeably, both referring to the point in time when a stop closure is first achieved. The same is true of 'end' and 'release', but we often use 'release' when the focus is on the acoustic and perceptual consequences of the end of consonant closure.

[8] As noted above, Biteeva (2021) found roughly simultaneous coordination of primary and secondary articulation in /tʲ/, unlike Kochetov (2009). In Biteeva's stimuli, the target coda preceded a word-initial coronal consonant, in which case burst cues might have been less reliable.

3.4 Hypotheses

Though there is not a great deal of previous evidence to go on, what evidence there is suggests that the relative timing of primary and secondary articulation gestures in palatalization contrasts is more variable in the coda than in the onset, across consonant types and speakers (and possibly within them as well). We would therefore expect to find such an onset-coda asymmetry in our Irish data.

If the guiding principle is that gestural timing is perceptually adaptive, then we expect both palatalization and velarization gestures to be aligned roughly with the release of onset stops in Irish, because that is where all cues to the contrast can be maximized. As noted above, previous work on Irish and Russian already leads us to expect this pattern for palatalized consonants. As for velarization, past work has *not* found any consistent tendency for /C$^{\gamma}$/ gestures to align with consonant release in onset position, as observed for /C^{j}/. Regardless of syllabic position, velarization seems fairly static over the duration of the consonant. However, relevant data are limited.

In the case of coda stops, where cues to the palatalization contrast are split between the beginning and end of the primary place gesture, we have no general reason to expect alignment specifically with release. Assuming both formant transition and burst cues are important, we would expect primary and secondary articulation gestures to be roughly simultaneous. Put differently, we might expect secondary articulation gestures to be achieved by the beginning of the consonant and maintained through to the release.

However, as discussed earlier, perhaps not all cues are equal in every context. If formant transitions are more effective cues than burst properties are for labial stops, then we might expect alignment of secondary articulations with the beginning of coda labials, at the vowel-consonant transition. If the reverse is true for coronal stops, we might expect alignment with the coda consonant release instead. However, these latter predictions are somewhat speculative.

4 Ultrasound study

4.1 Methods

4.2 Participants

We recorded 7 native Irish speakers, whose language background corresponded to each of the three major Irish dialect groups (Figure 2). The shaded portions

However, in Kochetov's stimuli, the target coda was prevocalic. An interesting question for future work is whether such differences are systematic.

of Figure 2 show the Gaeltachtaí - areas where Irish is spoken as a community language. It is conventional to distinguish three major dialect areas for Irish, that of Ulster in the northwest, Connacht in the west, and Munster in the southwest. Our study includes speakers from all three dialect areas, and we use these terms, but it should be noted that there is dialectal diversity within each dialect area.

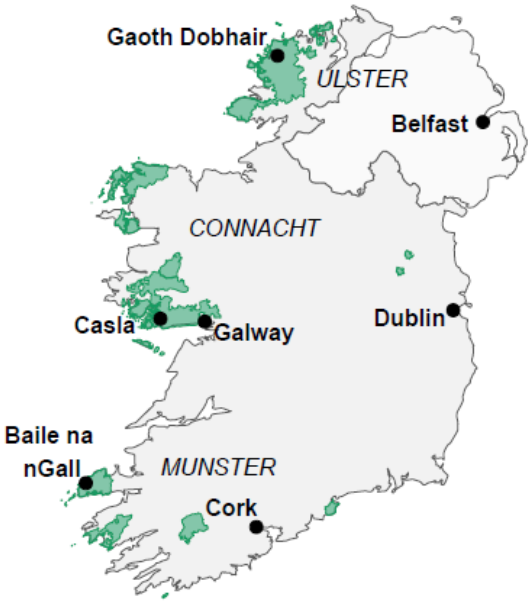


Figure 2: Officially recognized *gaeltachtaí* (Irish-speaking areas) in the Republic of Ireland, shaded in green (*gaeltachtaí* map data provided by `https://data-osi.opendata.arcgis.com/`, under CC by 4.0 license, `https://creativecommons.org/licenses/by/4.0/`).

We recorded two speakers of Ulster Irish (U1, 24, M; U2, 40, M), three of Connacht Irish (C1, 42, F; C2, 50, F; C3, 43, F), and two of Munster Irish (M1, 34, M; M2, 56, M). (Mean age = 41, median = 43.) Recordings took place in Gaoth Dobhair, Casla, and Baile na nGall, indicated on the map.

All of our speakers were professional Irish-language broadcasters, working for RTÉ Raidió na Gaeltachta, the Irish-language radio network of Raidió Teilifís Éireann, Ireland's national public service media system. Their jobs required them to speak and read Irish throughout the workday. All were native speakers of their respective dialects who were raised, and currently live, in the relevant dialect area. Their parents were native speakers who spoke Irish to them, except that one speaker's father (C3) was a native English speaker who spoke both English

and Irish to her, and another's (M1) mother was a native English speaker who spoke Irish to him. They attended primary and secondary school through Irish in their respective dialect areas, though C1-3 and M1 used both English and Irish in secondary school. All speakers attended third level colleges. All reported reading Irish on a daily basis. None reported any difficulties in hearing, speaking, or reading.

As in Bennett et al. (2018), we chose to work with Irish-language broadcasters for several reasons. First, even in traditional Irish-speaking communities (*gaeltachtaí*), English has grown increasingly dominant, particularly among younger speakers, as noted earlier. In this fragile sociolinguistic context, native speakers may avoid interactions – such as experimental phonetic studies – in which they perceive that their Irish is being evaluated. The speakers for this study represent relatively traditional varieties of Irish, are comfortable with audio recording technology, and are accustomed to speaking while being recorded.

Given the number of speakers of each dialect area, we cannot infer anything firm about dialect differences based on our study, though the speakers clearly showed expected properties of their dialects. (For example, the Ulster speakers pronounced /tʲ dʲ/ as [t͡ɕ d͡ʑ] respectively.) A tentative investigation of inter-speaker differences is provided in Section 7.

4.3 Items

The materials were (mostly) monosyllabic words with a target stop consonant controlled for secondary articulation (/Cʲ/ or /Cˠ/), place of articulation (labial, coronal, dorsal), position (word-initial or -final) and adjacent vowel /iː uː ɔː/ (= 2 × 3 × 2 × 3 = 36 total conditions). There was one word per condition. The words used are given in Tables 9 and 10 in Appendix A.

We used long /iː uː ɔː/ for the vowel contexts because these vowels vary in both backness – which is key to /Cˠ Cʲ/ contrasts in Irish – as well as height. Short vowels cannot be used for this purpose, because non-low short vowels take on the backness of a following consonant (Ní Chiosáin 1991; Ó Siadhail 1991): they are predictably front before palatalized consonants (e.g. *min* [mʲi̠nʲ] 'grain') and back before velarized consonants (e.g. [mʲu̠nˠ] 'small').

In using all three of /iː uː ɔː/, this study expands on Bennett et al. (2018), who only considered onset consonants preceding /iː uː/. Having said this, we found substantial variation in the pronunciation of /ɔː/ depending on dialect, speaker, context and word, particularly for the Ulster speakers, who had qualities including [ɔ a ɑ ʌ ɛ]. The high vowels were more consistent, though we saw some variation in the backness of /uː/ depending on the speaker and word. Due to these

and other unanticipated variations in vowel production, the conclusions we can draw about vowel context effects are limited.[9]

Since perfect control of materials would have made it impossible to use actual words, the target stops in our word list were either voiceless or voiced as needed (e.g. *tuí* /t̲ʸiː/ 'straw' vs. *díon* /d̲ʲiːnʸ/ 'roof'). For similar reasons, three of our forms are disyllabic instead of monosyllabic, though in all cases the target consonant was an onset/word-initial one in the stressed (initial) syllable (these are *píosa* /p̲ʲiːsʸə/ 'piece', *púca* /p̲ʸuːkʸə/ 'ghost', and *ciúnas* /k̲ʲuːnʸəsʸ/ 'quiet (noun)'). For one word, /bʲeːkʲ/ 'shout' with a final /kʲ/ target consonant, the vowel context was /eː/ instead of /iː/. In the case of onset/word-initial consonants, the nearest consonant (after the following vowel), if there was one, was a velarized coronal (e.g. /dʲiːnʸ/ 'roof'), except in the case of /pʸuːkʸə/ 'ghost'. In the case of coda/word-final target consonants, the nearest consonant (before the preceding vowel) was more varied (see Table 10 in Appendix A), again to ensure that actual words were used.

Items were recorded in the context of the carrier sentence [ˈdʸuːrʲtʲ ˈiːfʸə ____ əˈnʸʊrʸə] 'Aoife said ____ last year'. This means that word-final target consonants were positioned before a vowel-initial word, something that made segmentation and labeling based on the acoustic signal easier. However, it does raise the question whether the relevant target consonants are actually syllable-final in these sentences, as we assume they are. We do not believe that Irish has resyllabification between lexical words: we refer the reader to arguments in Green (2001) that resyllabification in Irish is possible only in the case of certain proclitics.[10] Note also that Kochetov (2006)'s finding of different timing for word-final labials compared to word-initial ones, discussed in Section 3.2, was also based on items placed in a carrier sentence before a vowel. As we will see, our own timing results also show that final stops do not behave like initial ones.

4.4 Recording procedure

Recordings were done in RTÉ Raidió na Gaeltachta studios in Gaoth Dobhair (Gweedore, Ulster), Casla (Costelloe, Connacht) and Baile na nGall (unofficially Ballydavid, Munster). Speakers repeated the 36 experimental items five times each in the frame sentence described above. All speakers read the words in the same pseudo-random order.

[9] In two forms the vowel context differed from the expected because of diphthongized pronunciations: [iːədʸ] for *iad* 'them' and [bʸuːɨkʲ] or [bʸuːɪkʲ] for *buaic* 'pinnacle'.

[10] For discussion of Irish syllabification in general, see Ní Chiosáin (Forthcoming).

4.4.1 Ultrasound and audio recording

We recorded midsagittal images of the tongue surface using a portable Terason T3000 ultrasound system and model 8MC3 3–8 MHz probe with a 90° field of view (using a depth setting of 10 and focal setting of 8, giving ≈ 46 fps, or 1 frame every 22 ms), using the Ultraspeech software package (Hueber et al. 2008), which automatically aligned the audio and video recordings. The probe was stabilized and held in place with an Articulate Instruments ultrasound stabilization headset (Wrench 2008; Scobbie et al. 2008). Simultaneous audio was recorded using a Shure WH20 dynamic cardioid microphone attached to the headset, recording directly to the ultrasound system (which included a laptop computer) at a 44.1 kHz sampling rate.[11]

4.5 Data annotation

4.5.1 Landmark identification and frame

Temporal landmarks for each consonant were annotated in Praat (Boersma & Weenink 2020), on the basis of acoustic information present in the waveform and spectrogram of the audio signal (see Turk et al. 2006). The beginning of C closure was primarily identified by a drop in amplitude from the preceding vowel. For onset stops, this vowel belonged to the preceding word in the frame sentence (the proper name *Aoife* [ˈiːfʲə]). For coda stops, the preceding vowel belonged to the same target word. Stop offsets were marked at the beginning of the stop release burst. In cases of uncertainty, the choice was generally made to err in the direction of smaller consonant duration, so as to minimize the possibility of including vocalic information as part of the consonant.

We extracted ultrasound frames for analysis at three different landmarks: stop onset, stop release, and the midpoint between them. (These correspond to the terms 'C start', 'C end', and 'C midpoint' used below.) These landmarks were determined as described above (see also Figure 1). Analyzing secondary articulations at each of these three landmarks should provide us with at least a rough measure of how those secondary articulations are timed relative to major articulatory events in the corresponding primary articulations of the same consonant.

[11]The microphone failed for the Connacht recordings, and the laptop's built-in microphone had to be used instead.

4.5.2 Contour tracing

Mid-sagittal ultrasound images were traced using EdgeTrak software (Li et al. 2005). EdgeTrak produces contours consisting of 100 points per traced tongue surface (Figure 3), each with (X,Y) coordinates in Cartesian space.

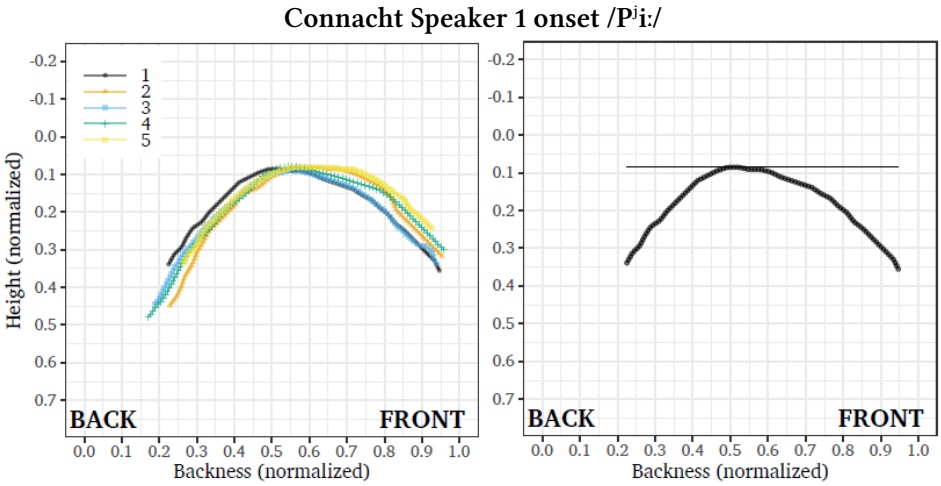


Figure 3: Left: sample EdgeTrack tracings for 5 tokens of onset /Pʲiː/ at C end, Connacht Speaker 1. Right: example identification of dorsal peak for one EdgeTrak tracing in this set (repetition 1).

Some mid-sagittal images could not be traced due to issues with image quality. This led to imbalances in the number of tokens across speakers and places of articulation. For some speakers, all tokens at a particular place of articulation could not be traced reliably, usually because the tongue surface for that consonant type consistently moved outside of the region in which the tongue body could be confidently identified (i.e. outside of the field of view and/or depth of the image produced by our ultrasound probe). This most often occurred for dorsal stops, and sometimes for coronal stops. More detailed information about the number of tokens traced per speaker is provided in Table 11 in Appendix B.

4.6 Quantitative analysis

4.6.1 Normalization

In order to ensure that contour tracings were comparable across speakers, all tracings were range-normalized on a by-speaker basis. The X-coordinates were rescaled to the interval $[0, 1]$, and Y-coordinates were rescaled by the same proportion and similarly shifted to have a minimum value of 0 (see Figure 3).

4.6.2 Dorsal peak backness

To assess the degree of velarization (= tongue body backing) and palatalization (= tongue body fronting) across contexts, we identified the X coordinate of the highest point of the tongue body in each contour (Figure 3). The position of this X coordinate can then be used as a measure of the degree of fronting or backing of the tongue body in a given context.

In some tokens, there was a plateau rather than a unique highest point for the traced contour. In such cases, the center of the plateau was selected as the 'peak' of the tongue body and the X coordinate of that center point used as our measure of tongue body fronting/backing.

We visually inspected each raw tracing to ensure that the highest point of the tongue contour was in a region that plausibly corresponded to the tongue body. For one speaker's coronal data (M2), the highest point of the tongue surface fell on the tongue blade or tip, and not the tongue body; this data has been excluded from the analysis of dorsal backness.

Overall, we examined the X position of the dorsal peak for 782 tokens in our data. The data and R scripts used to carry out this analysis are available on GitHub.[12]

4.6.3 Dependent measure: Change in backness over time

Our primary research questions concern the timing of secondary articulations in Irish. In our statistical analysis, we investigated differences in dorsal backness at C end vs. C start, for each token in our data. To do this, for each token we subtracted the dorsal backness value at C start from the corresponding value at C end. This difference – dorsal backness at C end relative to C start in each token – is the dependent variable for our analysis. Values that are different from zero indicate a difference in the magnitude of dorsal backing/fronting across time points; if consistent across tokens, this would suggest that the secondary articulation gestures for /C$^{\gamma}$ C^{j}/ peak at either C end or at C start in a given condition.

Such skews can be taken as evidence for a particular pattern of gestural alignment between primary and secondary articulations. For example, if dorsal fronting peaks at C end (= release) for /C^{j}/, we expect values of C end – C start to be positive for tokens of /C^{j}/, because the tongue should move forward over time. If dorsal backing for /C$^{\gamma}$/ also peaks at C release, we expect values of C end – C

[12] https://github.com/rbennett24/articles/tree/master/Irish_pal_timing_syll_FACL

start to be *negative* for /C$^{\gamma}$/, since velarization involves backing the tongue body rather than fronting it.

Somewhat more succinctly: positive numbers mean the tongue body moves forward over time during the consonant, whereas negative numbers mean the tongue body moves backward. Consistent patterns of either type can be used as evidence for timing relationships between primary and secondary articulations.

5 Results

Figure 4 shows dorsal backness trajectories, from C start to C midpoint to C end, for each token in our data. Certain tendencies are already apparent. For example, the position of the tongue dorsum appears to be further front at C release, compared to C start, for onset /P^{j} T^{j} K^{j}/, suggesting gestural alignment between the secondary /C^{j}/ articulation and stop release. Similarly, the tongue dorsum is further *back* at C release compared to C start for onset /P$^{\gamma}$/. We also observe a certain amount of variability in timing patterns across tokens in the same condition. Some of this variability may owe to the influence of vowel context (e.g. Bennett et al. 2024). Below we focus on other sources of variability.[13]

To help interpret the values presented in Figure 4 and subsequent figures, Tables 3 and 4 provide some summary statistics about the distribution of dorsal backness values in our data. For example, a difference of 0.11 between start and end position in Figure 4 would correspond to a difference of about one standard deviation for the full set of backness values in our data (across all three of C start, midpoint and end), or about 1/6 of the overall range of these values (Table 3). It would also be about the same size as the difference in mean dorsal backness between velar /K^{j}/ vs. /K$^{\gamma}$/ in onset position at C end in our data (Table 4). Figures 3 and 5 are also useful for visualizing these differences; see Table 5 for related statistical analysis.

We statistically analyzed values of tongue body backness at C end (relative to C start) using linear mixed-effects modeling in R (R Development Core Team 2020) with the lmerTest package (Kuznetsova et al. 2020; see also e.g. Pinheiro & Bates 2000; Gelman & Hill 2006; Baayen et al. 2008; Bolker et al. 2009; Barr et al. 2013; Matuschek et al. 2017; Tomaschek et al. 2018; Oberpriller et al. 2022). The

[13] A point of clarification: for each token, the value at C end is relative to C start. As a consequence, the values at C end should *not* be interpreted as reflecting the overall degree of palatalization (fronting) or velarization (backing) in a given token. Instead, values at C end should be interpreted as reflecting *change in backness over time*. (The same goes for values at C midpoint, of course, though we do not analyze that data here.) See Bennett et al. (2024) for an analysis of the strength of palatalization/velarization in this data.

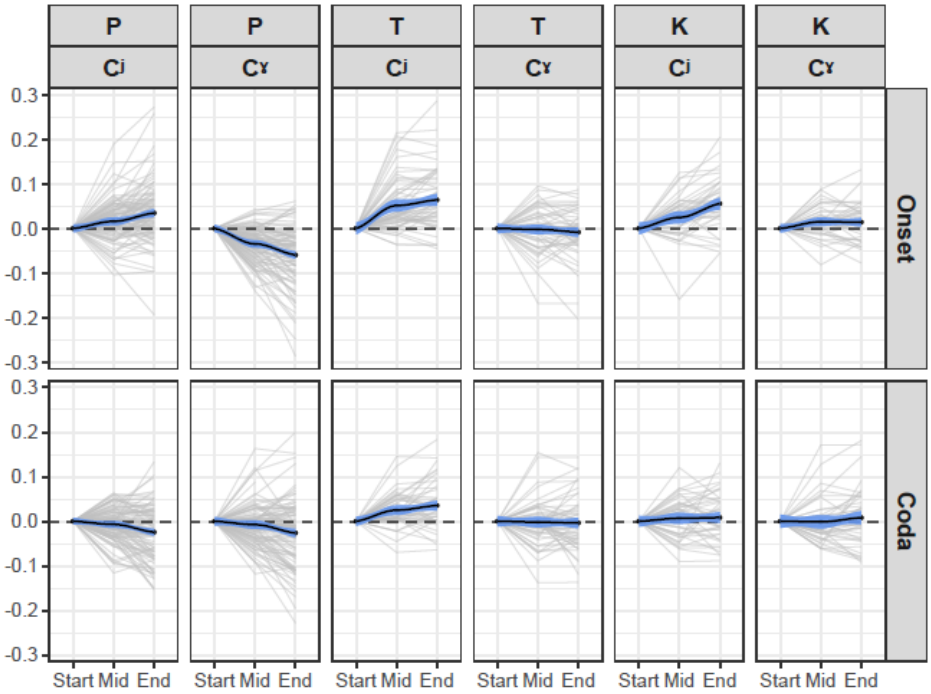


Figure 4: Trajectories for backness of dorsal peak over C start, midpoint, and end time points, relative to values at C start. Grey lines represent trajectories for individual tokens and black lines represent loess-smoothed regressions over pooled values, with an overlaid ribbon in blue showing confidence intervals around estimates of the mean.

Table 3: Summary statistics for distributions of dorsal backness values in our data. SD = standard deviation, IQ RANGE = inter-quartile range, Δ(X) = size of range for X.

Mean	Median	Range	Δ(Range)	SD	IQ RANGE	Δ(IQ RANGE)
0.5020	0.5080	[0.1356, 0.7789]	0.643	0.108	[0.4266, 0.5952]	0.173

predictors CONSONANT PLACE, SECONDARY ARTICULATION, SYLLABLE POSITION, and VOWEL CONTEXT were sum-coded in order to reduce collinearity between simple predictors and higher-order interactions, and to help with model interpretability. All two-way interactions between these predictors were included in the initial model; we also included a three-way interaction between CONSONANT PLACE, SECONDARY ARTICULATION, and SYLLABLE POSITION. The reference level

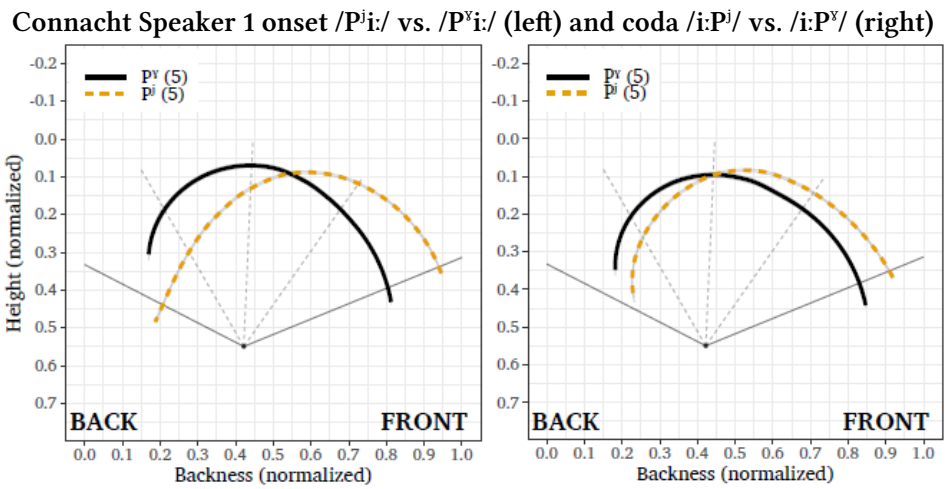


Figure 5: Loess-smoothed curves for /Pʲ/ vs. /Pˠ/ in the context of /i:/, in both onset (left) and coda (right) position. Connacht Speaker 1 (see Figure 3).

Table 4: Differences between mean backness values of /Cʲ/ - mean backness values of /Cˠ/ at C end, across places of articulation and syllable positions

Syllable position	/Pʲ/ - /Pˠ/	/Tʲ/ - /Tˠ/	/Kʲ/ - /Kˠ/	Overall
Onset	0.166	0.206	0.100	0.163
Coda	0.132	0.190	0.079	0.134

for Consonant Place was Labial; for Vowel Context it was /i:/; for Syllable Position it was Onset; and for Secondary Articulation it was /Cʲ/.

A random effect for Speaker was included, along with by-speaker random slopes for Syllable Position. Models with additional by-speaker random slopes failed to converge, in part because of the unbalanced nature of our data (see Appendix B). We did not include a random effect of Word because we only had one word per condition of interest. The specifications for the model are provided in (1), where Consonant Place=C Place, Vowel Context=V Con, Syllable Position=Syll Pos, and Secondary Articulation=Sec Art.

(1) Linear mixed-effects model specifications:
Dependent variable: Backness of dorsal peak at C end - C start ~
Three-way (1): C Place × Syll Pos × Sec Art +

Two-way (6): C Place × Syll Pos + C Place × V Con + C Place × Sec Art + V Context × Syll Pos + V Context × Sec Art + Sec Art × Syll Pos +
Simple (4): C Place + Syll Pos + V Con + Sec Art +
Random effects: (1+Syll Pos|Speaker)

This model was fit with the parameter REML = False so that we could use the log-likelihood test to carry out model criticism and simplification. However, model criticism did not lead to the exclusion of any fixed-effect predictors (with $\alpha = 0.05$), starting with the highest-order (three-way) interaction term, and retaining all simple predictors participating in interactions. As such, our final model was the same as our initial model (1). The residuals appeared approximately normally-distributed in the model. This model has only limited collinearity: the maximum variance inflation factor (VIF) was 1.34, as reported by the `check_collinearity()` function in the performance package (Lüdecke et al. 2021). VIF values below 5 are generally accepted to signal fairly low collinearity (e.g. Tomaschek et al. 2018).

The model results are provided in Tables 5 and 6. The *p*-values we report were estimated by the lmerTest package. Results shaded in gray indicate certain significant (or near-significant) predictors for the hypotheses we consider here: namely, that the timing of secondary palatalization and/or velarization gestures may be less stable in codas than in onsets, and less stable in labials than in consonants at other places of articulation (especially coronals).

While the results in Tables 5 and 6 imply that gestural alignment may be influenced by a number of factors, the precise patterns of alignment indicated by these results are a bit difficult to interpret, given the various overlapping interactions and main effects. Still, certain patterns do seem apparent. For example, while velarized /C$^{\text{ˠ}}$/ shows evidence of backing at C end, relative to C start (negative estimate for /C$^{\text{ˠ}}$/), this tendency is much weaker in coda position (roughly comparable positive effect for Coda × /C$^{\text{ˠ}}$/). This suggests that velarized consonants may tend toward release alignment in onsets, but not in codas (or at least, not as strongly in codas).

However, the significant effects of Dorsal × /C$^{\text{ˠ}}$/, Coronal × /C$^{\text{ˠ}}$/, and Coronal × Coda suggest that these tendencies may vary depending on place of articulation – in particular, labial /P$^{\text{ˠ}}$/ may be driving most of the onset vs. coda asymmetries for velarized /C$^{\text{ˠ}}$/, as implied by Figure 4. At a minimum, it seems that consonant place, secondary articulation, and syllable position play *some* role in determining patterns of gestural alignment in our data.

To clarify the structure of our results, Figure 6 plots the distributions of backness values, relative to C start, for all combinations of consonant place, sec-

Table 5: Main effects of linear mixed-effects model parameters for analysis of dorsal backness trajectories (values at C end, relative to C start).

Predictor	Estimate	SE	p
(Intercept)	0.0054	0.0049	0.31
C place			<0.001*
Coronal	0.0099	0.0031	<0.01 *
Dorsal	0.0144	0.0034	<0.001*
Syllable position			<0.46
Coda	−0.0051	0.0065	0.46
Secondary articulation			<0.001*
/C$^{\text{ɣ}}$/	−0.0207	0.0020	<0.001*
V context			<0.001*
/uː/	0.0023	0.0028	0.43
/ɔː/	0.0150	0.0028	<0.001*

Range of values: [-0.2870, 0.2858] (Δ = 0.573)
Mean = 0.0003, median = 0, sd = 0.0716
Reference levels: labial, /C$^{\text{j}}$/, onset, /iː/

ondary articulation, and syllable position in our data. (We abstract away from vowel context because it is not related to the specific hypotheses we focus on here.) For each cell, we assessed whether the distribution of dorsal backness at C end was significantly different from zero (i.e. different from C start, since C start was subtracted from all time points for each token). This was determined by t-tests with conservative Bonferroni-corrected significance values of $\alpha = 0.05/12 = 0.0041$. Distributions significantly different from zero in Figure 6 indicate patterns of gestural timing in which dorsal backness for secondary articulations peaks at either C end or C start (note that we do not analyze midpoint values at all here).[14] We label such distributions with Rel for 'alignment with C release' (= C end), and VC for 'alignment with VC transition in coda position' (= C start in coda position). These are the only two patterns of alignment that our statistical analysis uncovered.

Figure 6 shows that there are clear asymmetries across conditions with respect to the timing of dorsal gestures for /C$^{\text{ɣ}}$ C$^{\text{j}}$/. In onset position, the dorsal fronting gesture for /C$^{\text{j}}$/ peaks at C release; this is consistent across labial, coronal, and

[14]The significant results labeled in Figure 6 remain the same if significance is assessed using the bootstrap estimation method described in Section 7 rather than t-tests.

Table 6: Interactions of linear mixed-effects model parameters for analysis of dorsal backness trajectories (values at C end, relative to C start).

Predictor	Estimate	SE	p
C place × Syllable position			$< .075$
Coronal:Coda	0.0072	0.0032	$< .05^*$
Dorsal:Coda	-0.0058	0.0034	.09
Secondary articulation × C place			$< .01^*$
/C$^{\text{ɣ}}$/:Coronal	-0.0067	0.0029	$< .05^*$
/C$^{\text{ɣ}}$/:Dorsal	0.0099	0.0031	$< .01^*$
Secondary articulation × Syllable position			$< .001^*$
/C$^{\text{ɣ}}$/:Coda	0.0141	0.0020	$< .001^*$
C place × V context			$< .001^*$
Coronal:/uː/	-0.0138	0.0041	$< .001^*$
Dorsal:/uː/	-0.0018	0.0044	.69
Coronal:/ɔː/	0.0078	0.0040	.05
Dorsal:/ɔː/	0.0062	0.0043	.15
V context × Syllable position			$< .001^*$
/uː/:Coda	0.0012	0.0027	.64
/ɔː/:Coda	0.0234	0.0026	$< .001^*$
Secondary Articulation × V context			$< .01^*$
/C$^{\text{ɣ}}$/:/uː/	0.0009	0.0026	.74
/C$^{\text{ɣ}}$/:/ɔː/	0.0067	0.0026	$< .05^*$
Secondary articulation × C place × Syllable Position			$< .01^*$
/C$^{\text{ɣ}}$/:Coda:Coronal	-0.0049	0.0029	.09
/C$^{\text{ɣ}}$/:Coda:Dorsal	-0.0041	0.0031	.18

Range of values: [-0.2870, 0.2858] (Δ = 0.573)
Mean = 0.0003, median = 0, sd = 0.0716
Reference levels: labial, /C$^{\text{j}}$/, onset, /iː/

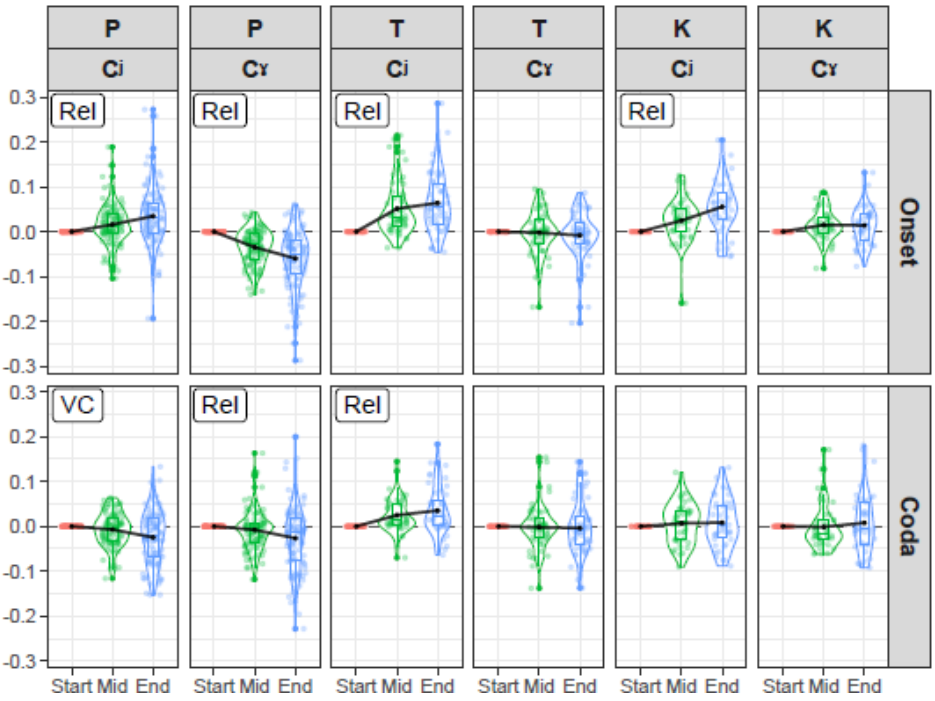


Figure 6: Values for backness of dorsal peak over C start, midpoint, end time points, relative to values at C start. Black lines connect mean values of distributions at each time point. Boxed text labels indicate statistically significant patterns of alignment in a given condition: 'Rel' indicates that alignment of the secondary articulation peaks at C release, and 'VC' indicates that alignment of the secondary articulation peaks at the VC transition in coda position.

dorsal places of articulation (see Figure 8 below for a visual illustration with onset /Pʲuː/). Additionally, velarized labial /Pˠ/ also shows release alignment in onset position. Velarized /Tˠ Kˠ/ show no particular pattern of alignment in onset position in this data.

In coda position, timing patterns are more varied. Coda /Pʲ/ shows VC alignment, while coda /Pˠ/ shows release alignment.[15] We illustrate this difference with a pair of examples from Ulster Speaker 1's data in Figure 7.

Both /Pˠ/ and /Pʲ/ show backing over time in coda position after /ɔː/ in Figure 7, indicating an *increase* in velarization over time, but a *decrease* in palatalization.

[15]It's important to remember that negative values at C end mean different things for /Cʲ/ vs. /Cˠ/. Negative values at C end indicate dorsal backing relative to C start: this implies stronger velarization over time, but *weaker* palatalization. So while the distributions of coda /Pʲ/ and /Pˠ/ are similar in Figure 6, they reflect different alignment patterns: release alignment for /Pˠ/ (stronger backing over time), and VC alignment for /Pʲ/ (weaker fronting over time).

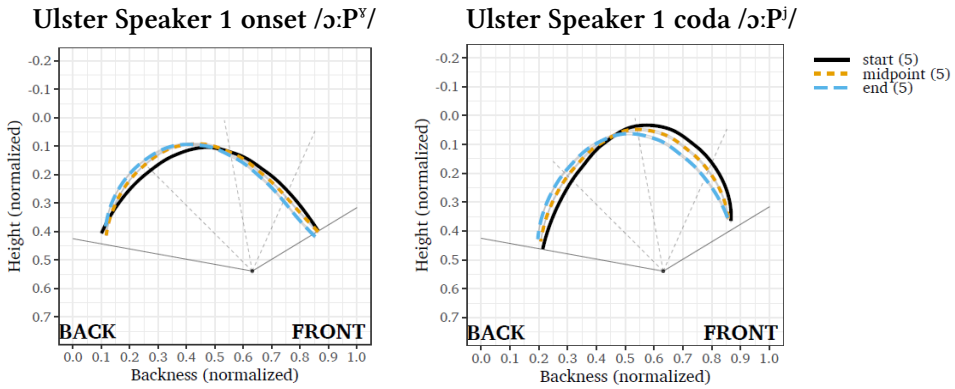


Figure 7: Loess curves for Ulster Speaker 1's coda /ɔːPˠ/ and /ɔːPʲ/ data showing change over time. The solid black line corresponds to C start, the dashed yellow line to C midpoint, and the dashed blue line to C end.

This reflects different timing patterns for coda /Pˠ Pʲ/: release alignment for coda /Pˠ/, and VC alignment for coda /Pʲ/.

Coda /Tʲ/ also shows release alignment in Figure 6, while the dorsal gesture for coda /Tˠ Kʲ Kˠ/ do not appear to show any particular alignment pattern at all.

Some of these results are also evident in the model output in Table 6. For example, the positive estimate for /Cˠ/ × Dorsal is consistent with the observation that coda labial /Pˠ/ (the reference level for C place) shows release alignment (negative values) across syllable positions, while dorsal /Kˠ/ shows values centered around zero for both onset and coda contexts.

The significant trends in Figure 6 are readily apparent from visual inspection of our raw tracings and corresponding loess-smoothed curves like Figure 7.[16] For example, Figure 8 confirms that the tongue body in onset /Pʲuː/ tends to be raised and/or fronted at C release (C end) relative to C start for most speakers. The degree of fronting involved varies – it is clearest for Connacht Speaker 2, and the Ulster speakers – but the overall pattern appears to hold for 5/7 participants (Connacht Speakers 1 and 3 being apparent exceptions).

Figure 9 provides a side-by-side comparison of raw tracings for Munster Speaker 1's onset /Pʲuː/ productions at C start and C release (C end). We note that (i) this speaker's productions are fairly consistent from token-to-token, and (ii) there is clear raising and fronting of the tongue body at C start relative to C

[16]The full set of raw tracings and loess curves discussed in this section is available at https://github.com/rbennett24/articles/tree/master/Irish_pal_timing_syll_FACL.

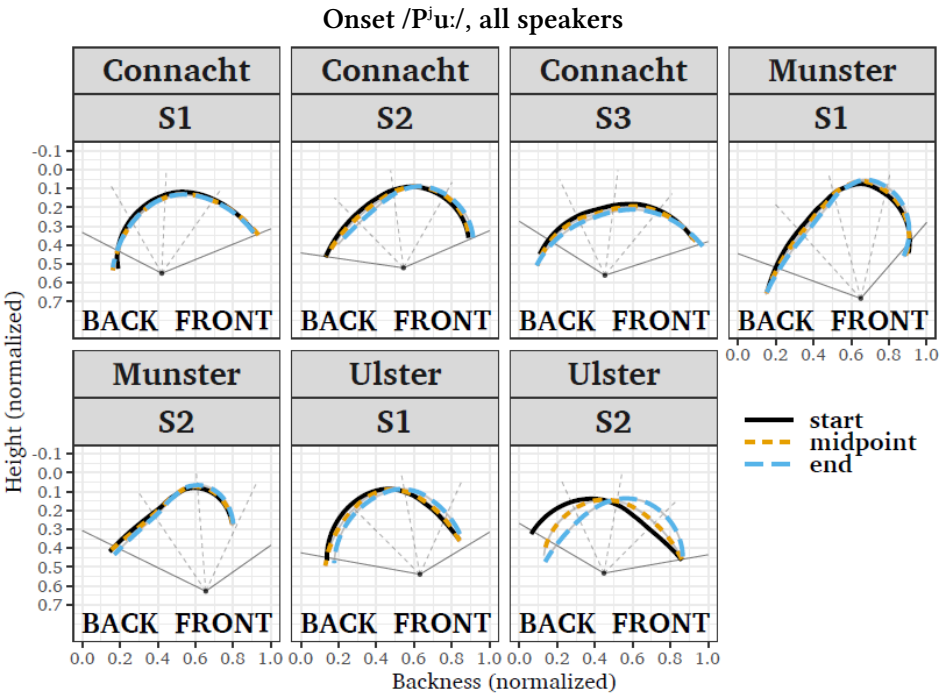


Figure 8: Loess-smoothed ultrasound tracings for every speaker's /Pʲuː/ tokens at C start, midpoint, and end.

end, despite the fact that the following vowel is back /uː/ (CV coarticulation for backness is limited in Irish; Ní Chasaide & Fealy 1991; Bennett et al. 2018; 2024).

We conclude that the significant results in in Figure 6 are reflected in qualitative patterns as well as quantitative ones.

The *non*-significant results in Figure 6 must be interpreted with caution, for two reasons. First, in frequentist statistics it is not generally valid to draw conclusions from the *lack* of a statistically-significant effect (this is sometimes called "accepting the null hypothesis"; Frick 1995 and many others).[17]

[17]In principle, Bayesian modeling can be used to justify acceptance of a null hypothesis (e.g. Kruschke 2013). However, in the specific case at hand, nothing is gained on this front by using Bayesian rather than frequentist statistics. Bayesian *t*-tests, implemented using the ttestBF() function in the BayesFactor package and the describe_posterior() function in the bayestestR package in R (Morey & Rouder 2023; Makowski et al. 2019), suggest that there is indeed sufficient evidence to accept the significant effects labeled in Figure 6. For the remaining conditions, the Bayesian *t*-tests suggest that the evidence is insufficient to either accept or reject the null hypothesis of no difference. When it comes to drawing inferences about the distributions in Figure 6, Bayesian methods leave us in the same place as regular frequentist ones.

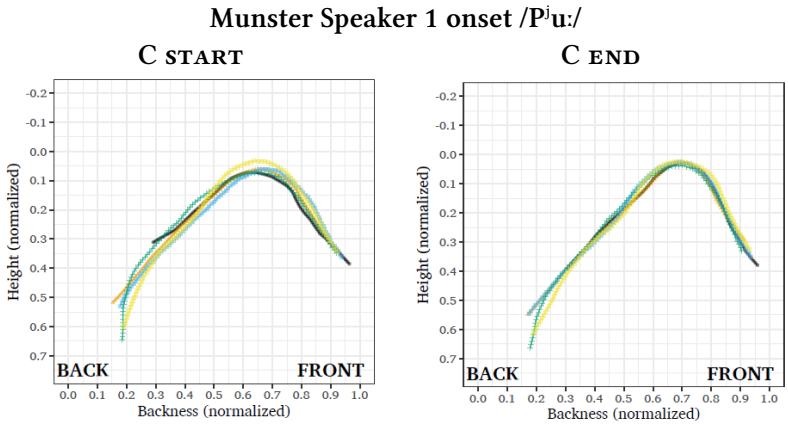


Figure 9: Raw ultrasound tracings for Munster Speaker 1's onset /Pʲuː/ tokens at C start (left) and end (right).

The second issue of interpretation surrounding Figure 6 is that a non-significant effect could correspond to two different scenarios regarding articulatory timing. First, it could be that backness values at C start are comparable to those at C end because the secondary articulation reaches its peak at C start, and is then maintained more or less constant until C end. In this scenario, we expect a fairly tight cluster of points around zero at C end in Figure 6, since the position of the tongue body should be more-or-less the same, consistently, at C start vs. C end.

Alternatively, it could be that articulatory timing is more noisy and variable: sometimes the secondary articulation peaks at C start, sometimes at C end, and sometimes in-between. In this scenario, we might expect a looser grouping of points around zero at C end. This is because differences in dorsal position at C end, relative to C start, would be less stable than in a scenario in which the secondary articulation peaks at C start and is held till C end.[18]

To try to disentangle these two interpretations, we can first consider some qualitative patterns in our data. Figure 10 provides loess-smoothed ultrasound

[18] A third scenario, which we reject, is that the peak of the secondary articulation varies categorically between C start and C end for those conditions that lack a significant skew away from zero. In this scenario, we would expect to find a *bimodal* distribution of points on either side of zero with one peak for tokens showing clear C start alignment and another peak, on the opposite side of zero, for tokens showing clear C end alignment. However, the distributions in Figure 6 appear to be consistently unimodal across conditions and time points, and so we reject this possibility. See also Padgett et al. (2023), who use Hartigan's dip test to argue that all of the distributions in Figure 6 are unimodal.

tracings for coda /Tˠ/ following /i:/ for one speaker (U2), at C start, midpoint, and end. We've chosen this context as an example because (i) coda /VTˠ/ is one of the non-significant conditions in Figure 6, and (ii) some of our other work on Irish suggests that velarized /Cˠ/ is often significantly fronted after /i:/ in coda position (Bennett et al. 2024). Thus, coda /i:Tˠ/ is a condition in which we might expect to find greater backing at C end than C start, where influence from the vowel /i:/ could pull the consonant articulation for /Tˠ/ forward at the VC transition.

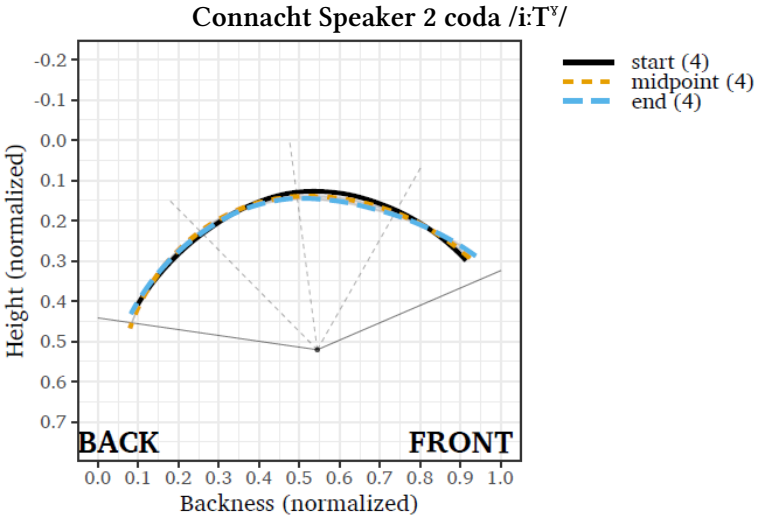


Figure 10: Loess-smoothed ultrasound tracings for Connacht Speaker 2's /i:Tˠ/ tokens at C start, midpoint, and end.

As is evident from Figure 10, tongue body position and posture are very similar across time points. There is perhaps a slight lowering of the tongue body at C end and midpoint relative to C start, but otherwise, there are very few differences when comparing these three landmarks. This implies that the constriction for /Tˠ/ is already achieved by the beginning of the consonant in /i:Tˠ/ for this speaker.

Loess-smoothed curves are a kind of average, which could potentially hide some informative variability across tokens. However, we can also see from Figure 11 that there is little variability in tongue body position across tokens at either C start or C end. This, too, is consistent with the claim that the constriction for /Tˠ/ is achieved at the VC transition (C start) and maintained basically constant until C release (C end).

Generally speaking, visual inspection of raw tracings and loess curves suggests that the non-significant conditions in Figure 6 resemble Figures 10 and 11: tongue body position is more or less the same at C start as at C end. Still, there

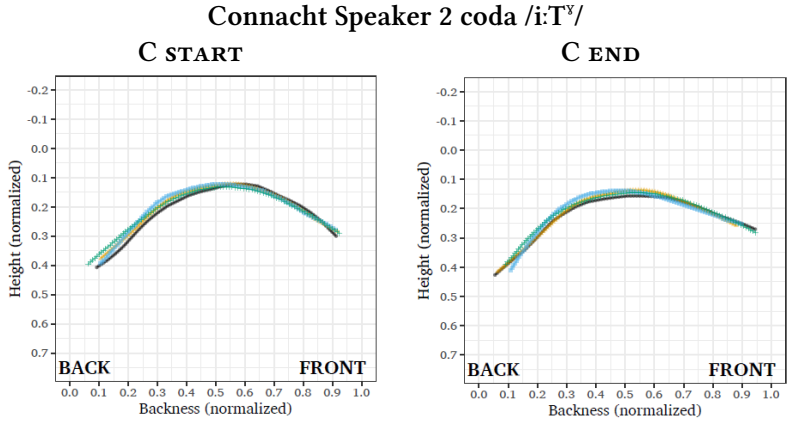


Figure 11: Raw ultrasound tracings for Connacht Speaker 2's coda /iːTˠ/ tokens at C start (left) and end (right).

are two interesting exceptions to this generalization. First, coda /Cˠ/ is sometimes realized with a fronter (less velarized) dorsal constriction at C start than at C end when following /iː/ (Figure 12). This reflects coarticulation for backness between /iː/ and a following /Cˠ/.

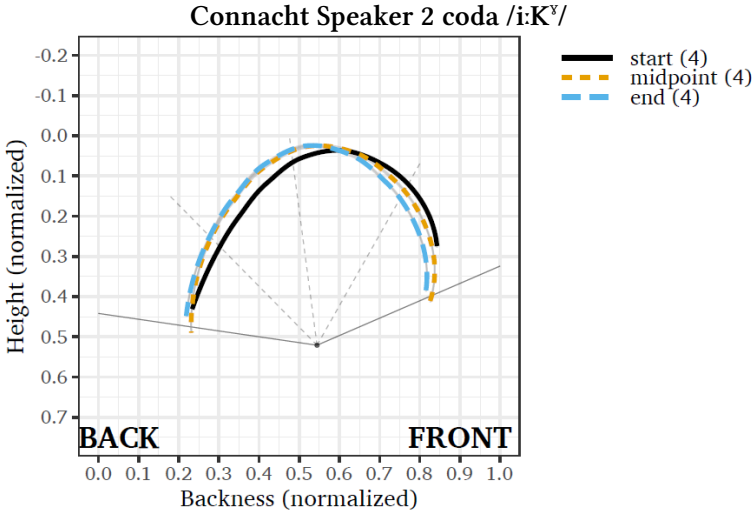


Figure 12: Loess-smoothed ultrasound tracings for Connacht Speaker 2's /iːKˠ/ tokens at C start, midpoint, and end.

At the VC transition, /Cˠ/ is fronter than at /Cˠ/ release, where the influence of the /iː/ is more removed. This is consistent with Bennett et al.'s (2024) find-

ing that /i:Cˠ/ is the context in which consonant backness is most affected by coarticulation with a neighboring vowel in this dataset.

Second, some speakers appear to transition towards a more neutral tongue body position at the end of coda C in some conditions. For example, Figure 13 shows that Ulster Speaker 2 has a fronter tongue body at C end than at C start in coda /u:Kˠ/. Both the vowel and consonant are back in this context, so fronting is not motivated by the segments themselves.

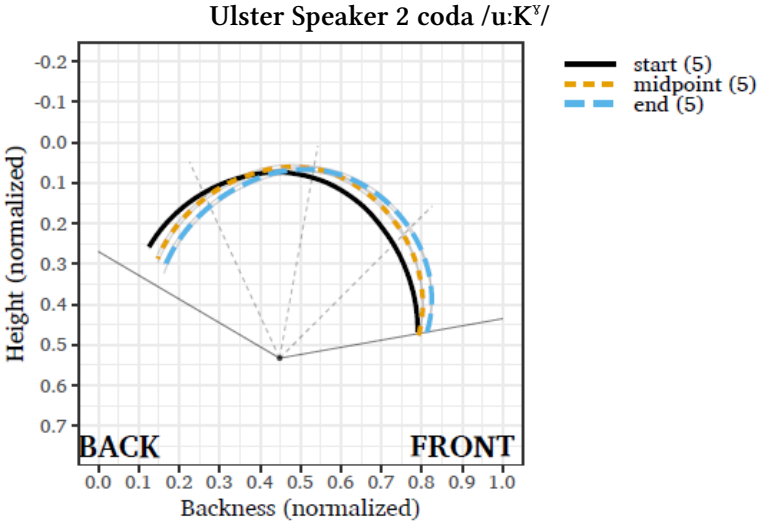


Figure 13: Loess-smoothed ultrasound tracings for Ulster Speaker 2's /u:Kˠ/ tokens at C start, midpoint, and end.

Recall that our target words were produced in a frame sentence (Section 4.3), which means that each target word was followed by the word [əˈnˠʊrˠə] 'last year'. The fronting observed in 13 could reflect anticipatory coarticulation with [ə] in [əˈnˠʊrˠə]. Alternatively, it could correspond to a phrase-final tongue posture of some kind, since our target words were generally produced with focus prosody, and so may have been prosodically phrase-final. (See e.g., Gick et al. 2004; Katsika 2016 for related discussion.)

Taking these observations together, our overall conclusion is that the non-significant conditions in Figure 6 generally represent a production pattern in which the secondary articulation is achieved at the VC transition (C start) and held more or less consonant through C release (C end). That said, we recognize that other factors (e.g. coarticulation) may play a role as well.

Further support for this interpretation comes from a more quantitative interpretation of the patterns in Figure 6. Our statistical question concerns the spread

of the distributions in Figure 6: in cases where backness values at C end overlap with those at C start, are the distributions at C end tightly clustered (≈ alignment at C start, continued to C end) or more widely spread out (≈ variable alignment)?

To answer this question, we can compare the variability in our non-significant conditions to the variability in conditions where we believe there *is* a specific, consistent pattern of gestural alignment. Table 7 reports summary statistics, including standard deviations and inter-quartile ranges, at C endpoint for each of the 12 conditions shown in Figure 6.

Table 7: Summary statistics for distributions shown at C endpoint in Figure 6. Starred rows indicate conditions with statistically significant differences from zero in Figure 6. SD = standard deviation, IQ RANGE = inter-quartile range, Δ(X) = size of range for X.

	MEAN	RANGE	Δ(RANGE)	SD	IQ RANGE	Δ(IQ RANGE)
*Onset P^j	0.017	[-0.194,0.272]	0.466	0.049	[0,0.036]	0.036
***Onset $P^ɣ$**	**-0.031**	**[-0.287,0.061]**	**0.348**	**0.050**	**[-0.056,0]**	**0.056**
***Onset T^j**	**0.038**	**[-0.045,0.286]**	**0.331**	**0.058**	**[0,0.062]**	**0.062**
Onset $T^ɣ$	-0.004	[-0.204,0.095]	0.299	0.041	[-0.011,0.004]	0.015
***Onset K^j**	**0.027**	**[-0.159,0.204]**	**0.364**	**0.049**	**[0,0.052]**	**0.052**
Onset $K^ɣ$	0.010	[-0.082,0.132]	0.214	0.032	[0,0.024]	0.024
*Coda P^j	-0.011	[-0.153,0.133]	0.286	0.042	[-0.027,0.004]	0.031
*Coda $P^ɣ$	-0.011	[-0.228,0.2]	0.428	0.051	[-0.027,0]	0.027
*Coda T^j	0.019	[-0.07,0.183]	0.253	0.040	[0,0.033]	0.033
Coda $T^ɣ$	-0.002	[-0.138,0.154]	0.292	0.044	[-0.016,0.003]	0.019
Coda K^j	0.005	[-0.09,0.131]	0.221	0.040	[-0.012,0.026]	0.038
Coda $K^ɣ$	0.002	[-0.092,0.181]	0.273	0.048	[-0.021,0.002]	0.022

As benchmarks, we can consider the three conditions with the largest deviations from zero, as determined by the absolute value of the mean. These are onset /$P^ɣ$ T^j K^j/, which are highlighted in gray and in bold in Table 7. These three conditions are the clearest examples of consistent articulatory alignment in our dataset, with secondary articulations seeming to peak at C end (= release) for all three conditions.

If we compare the measures of variance for the *non*-significant conditions in Figure 6 and Table 7, we see that those measures are similar in size to the measures of variance in our three significant benchmark conditions. For example, the mean of onset /$T^ɣ$/ (-0.004) is not significantly different from zero. The measures of dispersion for this condition – SD = 0.041, ΔRANGE = 0.299, and ΔIQ RANGE = 0.015 – are in fact *smaller* than the same measures for all three of our benchmark significant conditions (onset /$P^ɣ$ T^j K^j/). This suggests that the values for onset

/T$^{\gamma}$/ are at least as tightly clustered as the values for our significant benchmarks. The same general point can be made for *all* the non-significant conditions in Figure 6 and Table 7: the dispersion around the mean appears to be basically the same, and sometimes smaller, than found for the significant conditions demonstrating specific patterns of articulatory alignment. This is also consistent with simple visual inspection of the distributions in Figure 6.[19]

We conclude that the distributions for the non-significant conditions in Figure 6 and Table 7 are not particularly loose or spread-out around zero. Instead, they appear to be about as tightly clustered as the distributions in our significant conditions. This is consistent with the claim that the secondary articulations in our non-significant conditions typically reach a peak at C start, and are then maintained more-or-less unchanged until C end. In the discussion below, we sometimes use the term 'held alignment' to refer to this scenario, in which there is no significant change in dorsal backness from C start to C end.

6 Interim discussion

Recall our hypotheses from Section 3.4:

(1) The relative timing of primary and secondary articulations should be more variable in the coda than in the onset, across consonant types and speakers (and possibly within them as well).

Our measures of dispersion in Table 7 do not support this prediction. (Nor do they support an analogous hypothesis concerning labials compared to other places of articulation.) This hypothesis would lead us to expect more spread-out distributions for codas, relative to onsets, at C end in Figure 6. This is not the case, either visually or numerically (Table 7). However, we find some *tentative* support for this hypothesis below when we examine the results by individual participants.

(2) In onset stops, both palatalization and velarization gestures should be aligned roughly with the release of the consonant closure.

[19] Padgett et al. (2023) report that the distribution for non-significant onset /T$^{\gamma}$/ is more spread out than expected for a normal distribution according to a Shapiro-Wilks test with Bonferroni-corrected $\alpha = 0.015/12 = 0.0042$. However, this is also true of significant onset /P^{j} P$^{\gamma}$ T^{j}/ and coda /T^{j}/, and so this measure of dispersion also fails to distinguish our significant and non-significant conditions in Figure 6.

As expected based on previous findings, for palatalization this prediction holds across all three places of articulation (Figures 4, 6): values at C end for onset /P^{j} T^{j} K^{j}/ skew positive, and are significantly different from zero, indicating greater fronting at C end relative to C start.

As we noted earlier, previous work has not found such a clear pattern for onset velarization, and we do not find compelling evidence *overall* for release alignment of velarization (= significant negative skew at C end) in this study either (but see the discussion below).

(3) In coda stops, both palatalization and velarization gestures should be roughly simultaneous with the primary articulation, achieved around the beginning of the consonant and maintained through to the release.

Here we must be more cautious, given limitations of our analysis discussed in the previous section. Still, our results support such an interpretation for /T^{ɣ} K^{j} K^{ɣ}/ (Figures 4, 6, discussion around Table 7): values at C end seem to be centered around zero for coda /T^{ɣ} K^{j} K^{ɣ}/.

Also in a more speculative vein, we hypothesized that secondary articulations in the coda might be more aligned with the *beginning* of labial closures but with the *end* of coronal closures. This is based on the relative strength of formant cues for labial stops vs. release cues for coronal stops, and it has been observed for limited Russian data (Section 3.3). Our findings depart from predictions 1 and 2 above in ways that are suggestive in this regard.

First, in our data neither secondary articulations in the coda nor velarization anywhere generally favors alignment to a particular landmark – C beginning or release – /P/ contradicts these trends in having alignment with a landmark in all four conditions (syllable position x secondary articulation). In three out of these four conditions alignment favors vowel formant transitions (release alignment for onsets, beginning or VC alignment in codas). The exception is coda /P^{ɣ}/, which favors release alignment. Second, coda /T^{j}/ also contradicts the above trends, in favoring release alignment. All of these facts except for that of coda /P^{ɣ}/ are consistent with the hypothesis that alignment of secondary articulations is fine-tuned not just to the location but to the relative *importance* of cues in labials vs. coronals. Assuming there is validity to these considerations, it's not clear why /P^{ɣ}/ does not follow the expectation.

There is a separate, articulatory, consideration relevant to the facts about labials. In the case of /T^{j} T^{ɣ}/ or /K^{j} K^{ɣ}/, there is tongue coupling, and possibly direct articulatory competition, between the primary articulator and a secondary dorsal one, while this is not true of /P^{j} P^{ɣ}/. Perhaps this allows other factors – such

as constraints on perceptibility like those discussed here – to more freely determine timing in the latter case. However, we *do* find alignment with C release in onset /T^{j} K^{j}/ and coda /T^{j}/, casting some doubt on the importance of such an articulatory effect. (See Bennett et al. 2018 for related discussion.)

7 Inter-speaker variation

In this section we provide a preliminary investigation of inter-speaker variation in timing patterns in our data. The goal is to see how well the generalizations above hold up by speaker. In addition, previous research has reported inter-speaker variation in the production of /C$^{\gamma}$ C^{j}/ contrasts in Scottish Gaelic (Sung et al. 2018) and in Russian (Kochetov 2002; 2009). As for Irish, drawing on the same dataset analyzed here, Bennett et al. (2024) find that speakers differ in how robustly /C$^{\gamma}$ C^{j}/ contrasts are distinguished by dorsal position across contexts. For example, some speakers seem to completely merge coda labial /P$^{\gamma}$ P^{j}/ following /i:/, at least in terms of dorsal shape and position, while other speakers maintain distinct dorsal postures for these consonants in this context.[20] This pattern of inter-speaker variation, along with others observed by Bennett et al. (2024), raises the question of whether speakers might differ with respect to the patterns of gestural timing reported in aggregate in Section 5.

Figure 14 shows loess-smoothed regressions for the trajectories of dorsal backness values across consonant start, midpoint, and end, for all combinations of syllable position, consonant place, and secondary articulation. This is akin to Figure 4, but here separate regressions are provided for each speaker. For the sake of readability, we omit lines corresponding to the trajectories of specific tokens, as well as confidence intervals around the loess-smoothed estimates.

In general, the gestural timing patterns we observed in Section 5 are broadly reflected in the patterns of timing observed for each individual's data. For example, the secondary articulation for onset /K^{j}/ seems to peak at C release for all three speakers whose data is included in this condition. Similarly, for both onset and coda /T^{j}/, speakers mostly show a consistent pattern of aligning peak dorsal fronting with C release. However, still focusing on /T^{j}/, speakers also seem to differ in how distinct dorsal position is at C end vs. C start for this sound, and two speakers seem to show no significant fronting over time for onset /T^{j}/.

[20] Bennett et al. (2024) report measures taken at C release for onset consonants, and at C start in the VC transition for coda consonants; they do not investigate possible changes in articulatory posture across time points within a given consonant, as we do here.

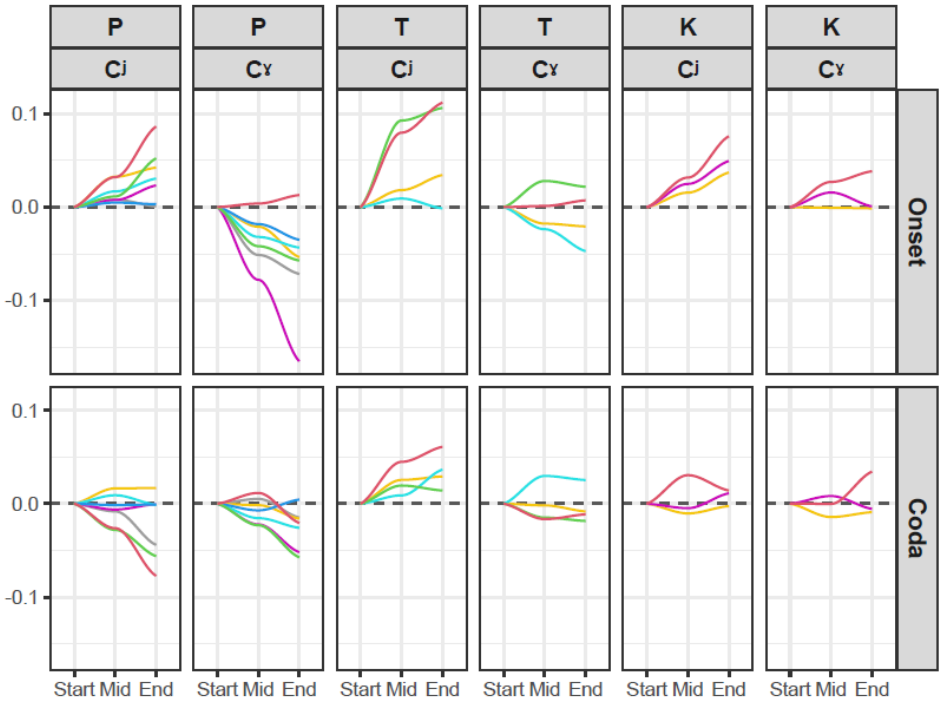


Figure 14: By-speaker trajectories for backness of dorsal peak over C start, midpoint, end time points, relative to values at C start. Colored lines represent loess-smoothed regressions over each speaker's values at each time point.

To quantify these patterns, we computed confidence intervals around the mean of each speaker's dorsal backness values at consonant end, relative to consonant start, using a non-parametric bootstrapping resampling method, using the SMEAN.CL.BOOT function in the HMISC package in R (Harrell Jr. 2023; R Development Core Team 2020). If the confidence interval around each speaker's mean at C end is significantly different from zero, this would indicate that the dorsal position of the secondary articulation peaks at release, assuming that the mean values are fronter for /Cʲ/ and backer for /Cˠ/. If instead the mean values are *backer* for /Cʲ/ and *fronter* for /Cˠ/, this would indicate that secondary articulations peak at C start.

In order to reduce the number of statistical comparisons involved, we limited this analysis to backness values at C end. We chose non-parametric bootstrapping to estimate the means and confidence intervals, rather than a parametric method like a *t*-test, because (i) bootstrapping methods do not assume that the

underlying data is normally distributed, and (ii) bootstrapping is more robust for smaller sample sizes. Sample size is a potential concern here, since splitting our data up by speaker substantially reduces the number of observations per comparison: for most combinations of speaker and condition, we have just 10–15 observations at C end (minimum = 8; 58/60 comparisons have at least 10 observations). In applying non-parametric bootstrapping to this data, we sampled with replacement 10,000 times for each combination of speaker and condition. This resulted in 60 total bootstrapped estimates (Figure 14).

We reiterate that caution must be taken in interpreting these results given the small sample sizes involved (see e.g. Chernick 2011: Ch.9 with respect to bootstrap methods, specifically). Nonetheless, we think it is useful to consider, in a broad and exploratory sense, the extent to which individual speakers might differ in their patterns of gestural timing for secondary articulations across contexts. Table 8 reports the significant results produced by this method with a significance threshold of $\alpha = 0.05$.[21]

Table 8: Significant differences between distributions of backness values, relative to C start, at C end, by condition. Counts refer to number of speakers showing each pattern.

	P^j	T^j	K^j
Onset	Release: 5/7	Release: 3/4	Release: 3/3
Coda	VC: 3/7	Release: 3/4	0/3 significant
	$P^ɣ$	$T^ɣ$	$K^ɣ$
Onset	Release: 6/7	Release: 2/4, C start: 1/4	C start: 1/3
Coda	Release: 3/7	Release: 1/4	VC: 1/3

Again, the results in Table 8 broadly reflect the findings of Section 5. First, palatalized /C^j/ consistently shows release alignment in onset position. For each place of articulation, the majority of speakers represented in that condition show release alignment, though not everyone. Two speakers do not align the peak of onset palatalization with the release of /P^j/, and likewise for /T^j/. In most other

[21]Given the fact that Table 8 reports multiple comparisons, we should arguably use a Bonferroni-corrected significance threshold of $\alpha = 0.05/60 \approx 0.0008$. Because this is an extremely conservative threshold for significance, and because our aims here are mostly exploratory, we opted for the standard $\alpha = 0.05$ instead. Using the conservative threshold of $\alpha = 0.0008$ shrinks the number of significant results in Table 8 from 33 to 21, but does not affect the broad qualitative conclusions we would draw from this data.

conditions, few or no speakers show alignment with either consonant landmark. The exceptions are familiar from the discussion in Section 6: several speakers *do* align secondary articulations for labials and for /T^{j}/.

Figure 15 plots the same data as in Figure 14, but pooled across places of articulation, so that each estimated trajectory is based on a larger number of observations. One can see in this figure our overall findings that onset palatalization aligns with C release while coda secondary articulations remain relatively unchanged throughout. There is release alignment for onset velarization as well, but this is mostly attributable to onset /P^{ɣ}/ as noted previously.

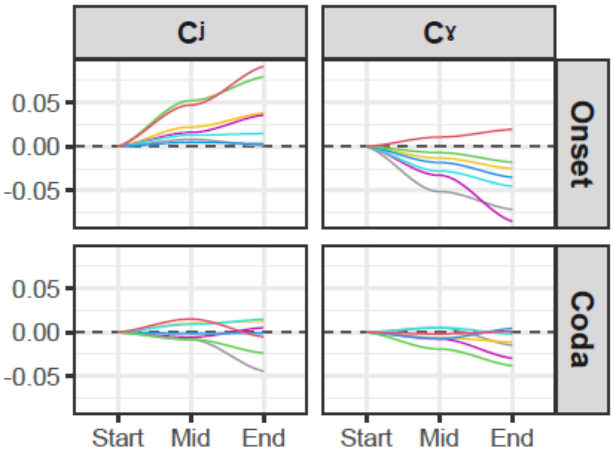


Figure 15: By-speaker trajectories for backness of dorsal peak over C start, midpoint, end time points, relative to values at C start. Colored lines represent loess-smoothed regressions over each speaker's values at each time point, collapsing across place of articulation.

On the whole, we can draw two tentative conclusions from the above. First, alignment patterns for individual speakers more or less replicate the overall population trends reported in Section 5. However, second, there are signs of the variability predicted by our first hypothesis in Section 6. Though we did not find greater variability for codas (or for labials) in our pooled data, here more *patterns* of alignment occur in codas than in onsets. Specifically, while *release* alignment is the only pattern found for onset consonants (apart from limited cases of C start alignment for velarized /T^{ɣ} P^{ɣ}/), there is release, VC transition, and held alignment in the case of codas. [22] There is likewise some suggestive evidence

[22]When alignment is not aligned uniquely with the beginning or end of the consonant (= 'held alignment'), this might be understood as alignment with both, i.e. simultaneous production, as discussed earlier.

here that speakers may differ, to some extent, in their preferred timing patterns across different contexts.

To be sure, more work needs to be done to verify that these patterns of variability reflect bonafide inter-speaker differences, rather than accidental facts about our particular sample of speakers and recordings. Confirming these patterns of variability in a larger dataset and exploring potential underlying causes (e.g. dialect variation) are clear topics for future research.

8 Against anti-phase timing as the source of onset-coda asymmetries in palatalization

Our discussion has assumed that secondary dorsal articulations in Irish are coordinated with consonantal landmarks, as defined by the production of the consonant's primary articulation. This strikes us as a reasonable assumption: after all, palatalization and velarization are properties of consonants, at least phonologically.

Beyond alignment with consonant landmarks, we have suggested (Section 3.3) that timing of secondary articulations might be constrained in perceptually adaptive ways. For example, in coda position, speakers must 'choose' whether to align the peak of a secondary articulation near the V-C boundary, where formant transition cues reside, or at C release, where there are cues from burst quality, or with both landmarks. Both of these landmarks are important, perceptually speaking, for expressing secondary articulation contrasts. In coda position, unlike in onset position, these landmarks do not coincide, and so speakers can align peak /C^{j} C^{γ}/ gestures at the V-C transition, or at the C release; or they can try to maintain the peak /C^{j} C^{γ}/ gesture throughout the entire consonant, so that the gesture occurs at *both* the V-C transition and C release. Evidently, speakers adopt different timing strategies in coda position depending on the place and secondary articulation of the consonant, in ways that seem largely consistent with the hypothesis of perceptual optimization (Section 5).

However, given the broad asymmetry we find between patterns of gestural alignment in the onset vs. coda for palatalization, we might look elsewhere for an understanding of what is happening. Studies of articulatory coordination have reported asymmetries in how segment-internal gestures are timed in onset vs. coda position. To illustrate, consider the American English lateral /l/, which has both a primary tongue tip constriction, and a secondary, vowel-like dorsal constriction (e.g. Krakow 1999; Sproat & Fujimura 1993; Proctor 2009; Lee-Kim et al. 2013; Turton 2017). In onset position (e.g. *l̪ab*), the coronal and dorsal gestures

for /l/ are largely synchronous. However, in coda position (e.g. *ball*), they appear to be produced sequentially, with the dorsal constriction preceding the coronal constriction in time. In codas, the overlap of the dorsal constriction for /l/ with the preceding vowel leads to a distinct and highly audible coloration of the vowel itself.

Similar observations have been made for the relative timing of oral closure and velum lowering in the American English nasals /m n/ (Krakow 1999; Byrd et al. 2009). In this case, velar lowering is largely synchronous with oral closure formation in onsets, but precedes closure formation in codas, producing extensive coarticulatory nasalization on the preceding vowel. Generalizing over these results, we can perhaps say that more open constrictions tend to precede more narrow constrictions for articulatorily complex consonants in coda position, at least in American English (Krakow 1999; Sproat & Fujimura 1993; Iskarous & Kavitskaya 2010).

These asymmetries in articulatory timing have been interpreted as a reflection of different underlying patterns of gestural coordination in onsets vs. codas. In onsets, gestures tend to be coordinated *in-phase*, or synchronously. In codas, gestures tend to be coordinated *anti-phase*, or sequentially. It has been argued that in-phase coordination is inherently more stable than anti-phase coordination, not just in speech but in motor planning and execution more generally (e.g. Goldstein et al. 2006; Nam et al. 2009; Parrell 2012 and references there). We might thus expect gestural alignment between primary and secondary constrictions to be more variable in codas for /C^{j} C^{ɣ}/, exactly as we find (at least for palatalization), if those constrictions are timed in an anti-phase relationship.

There are at least two reasons to doubt this alternative analysis. First, Irish palatalization timing is more variable in the coda only in the sense of allowing more *patterns* of alignment depending on place of articulation (and to some extent speaker), as just discussed: alignment tends toward C start for /P^{j}/, C release for /T^{j}/, and neither for /K^{j}/. In other words, as shown in Figure 6 and Table 7, differences between onset and coda position in our study mostly boil down to with which landmark the secondary articulation is coordinated. Measures of variance for the timing of primary and secondary articulations actually imply similar variance in onset vs. coda position, *not* increased variance in codas.

Second, palatalization is *not* simultaneous with the primary constriction in onset position; rather, this study and previous ones have found a very consistent lag in the palatalization gesture, in both Irish and Russian, such that the peak occurs around the release of the primary constriction (see also Shaw et al. 2021). As for coda position, though secondary articulations in our data peak around the onset of the primary constriction in the case of /P^{j}/, this is not true at all for other

conditions. Furthermore, as we saw in Section 3.3, for Russian consonants other than stops, the facts suggest secondary articulations that are co-extensive with the primary ones, as we have found for some conditions here.

In the current study, we can investigate this question empirically, building on the discussion of Russian in Iskarous & Kavitskaya (2010). If secondary /C^{j} C^{ɣ}/ are timed anti-phase (sequentially), then the gestural peak for /C^{j} C^{ɣ}/ should occur during the vowel, well prior to the achievement of the primary constriction. This would mirror findings for American English /l/ and nasals in coda position, in which the dorsal constriction for /l/ and the velar lowering for /m n/ are timed to reach a peak during the vowel itself.

A major acoustic correlate for the /C^{j} C^{ɣ}/ contrast in Irish (and other languages) is the F2 transition on neighboring vowels. F2 largely reflects dorsal backness in this context; it is also affected by lip rounding (e.g. Bennett et al. 2019). If /C^{j} C^{ɣ}/ gestures peak during the vowel in VC# coda sequences, as predicted by anti-phase alignment, F2 values should show evidence of peaking during the vowel as well.

To assess this prediction, we measured F2 during the vowel in /VC#/ sequences, where our target consonant was a word-final coda consonant. The vowel was segmented from the preceding consonant at the beginning of voicing: since our focus is on VC transitions, not CV transitions, the criterion used to identify the beginning of the vowel is not especially important. The vowel was segmented from the following target stop consonant at the point where spectral energy in the vowel abruptly dropped off, particularly in the region of F2 (see e.g. Turk et al. 2006). Any preaspiration on the following final consonant was segmented as part of the vowel itself, because significant formant transitions may occur during periods of aspiration (e.g., Stevens & Klatt 1974; see also Clayton 2010). Formant tracking may be less accurate during preaspiration than during vowels, given the presence of aperiodic noise and the relative weakness of higher formants during aspiration. That said, visual inspection of a number of preaspirated tokens suggests that F2 was still tracked fairly well during preaspirated transitions. 630 vowels were segmented in this analysis.[23]

F1, F2, and F3 were extracted from each vowel using the Fast Track Praat plugin (Boersma & Weenink 2020; Barreda 2021).[24] The first three formants were measured every 2ms during the vowel, and the resulting formant trajectories were then time-normalized before plotting (Figure 16). Formants were extracted

[23]F1 tracking was noisy, particularly for our Ulster speakers. We assume this is due to preaspiration, which appears to have less of an effect on F2 tracking (the formant we're mostly interested in here).

[24]https://github.com/santiagobarreda/FastTrack

using the default settings of the Fast Track plugin, except that the number of analysis steps was set to 24.

In order to pool F2 measurements across speakers, each speaker's formant values were normalized using Barreda-Neary log-additive regression normalization (Barreda & Nearey 2018). These normalized values were highly correlated ($r = 0.96$) with normalized values produced by a different method, namely dividing F1 and F2 by F3 at the same measurement point (e.g. Monahan & Idsardi 2010). Outliers were excluded from analysis by removing values for normalized F2 that were greater than 2 standard deviations away from the mean of the normalized F2 values. This resulted in the elimination of 460 F2 measurements, about 0.7% of the original F2 data. This left 66,677 F2 measurements for analysis (630 vowels, with a mean of 106 F2 measurements per vowel).

Figure 16 provides loess-smoothed F2 trajectories during the vowel in /VC#/ sequences, comparing coda /C$^{\gamma}$ C^{j}/ at each place of articulation in combination with the vowels /iː ɔː uː/.

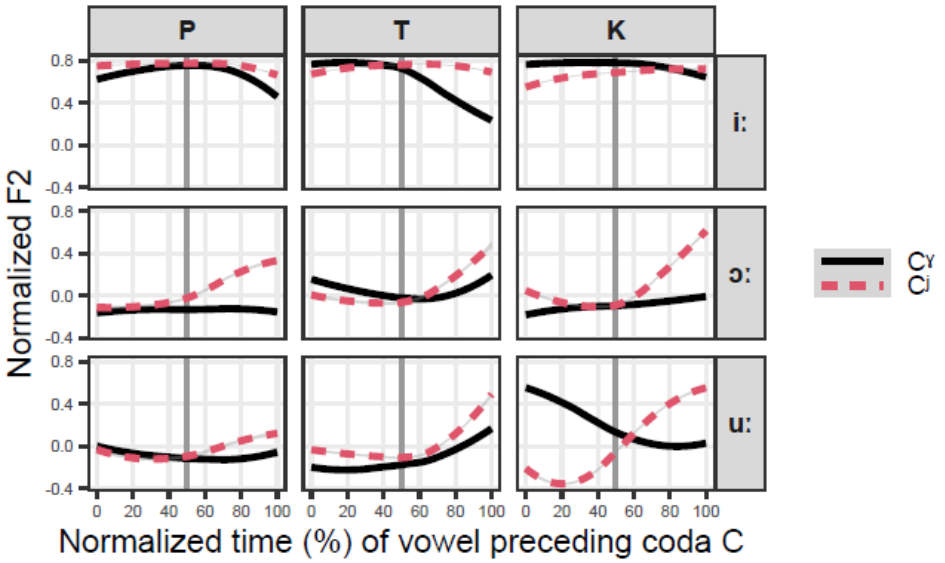


Figure 16: Pooled, speaker-normalized F2 trajectories across time-normalized steps, grouped by primary place of articulation and vowel context, for coda /VC#/. Lines represent loess-smoothed regressions over F2 values preceding /C$^{\gamma}$ C^{j}/ in each condition: a solid black line for /C$^{\gamma}$/, a dashed red line for /C^{j}/. A gray vertical line marks 50% point of the vowel.

Our empirical question is when F2 peaks for /C^{j}/, or reaches a minimum for /C^{ɣ}/, in /VC#/ sequences. Since F2 is our acoustic surrogate for dorsal fronting and backing, extreme values for F2 should correspond to the time points at which the tongue body is furthest front for /C^{j}/, or back for /C^{ɣ}/.

Answering this question is easiest for /VC#/ sequences in which the vowel and following consonant have different backness specifications. This is because we expect to find very salient F2 movements in cases of mismatched vowel and consonant backness (Bennett et al. 2018), which should make it easy to see where F2 trajectories reach their most extreme values. In our data, this corresponds to /iːC^{ɣ}/ sequences for velarized coda consonants, and /uːC^{j}/ and /ɔːC^{j}/ sequences for palatalized coda consonants.[25]

We begin with /uːC^{j}/, corresponding to dashed red lines in the bottom row of Figure 16. Here, it is clear that F2 begins to rise around V midpoint, or even earlier, and does not reach its peak until the beginning of the following consonant. The same pattern holds, perhaps even more clearly, for /C^{j}/ after /ɔː/ (dashed red line, middle row).

For /iːC^{ɣ}/ sequences (solid black line, top row) it is again clear that F2 reaches an extreme value at C start, rather than any earlier time point. A potential exception is /iːk^{ɣ}/, which shows a relatively flat F2 trajectory throughout the vowel; but even here we can see a slight lowering of F2 at the VC transition, into the beginning of the consonant.

The same patterns are visible if we plot F2 trajectories in physical rather than normalized time. Figure 17 shows vowel F2 for up to 200ms preceding the following coda consonant. Vertical arrows mark when F2 reaches its expected extreme value in each loess curve for /iːC^{ɣ} uːC^{j} ɔːC^{j}/, the contexts in which there is a mismatch in backness between the vowel and following coda consonant.

We can compare the timecourse of F2 movement in Figure 17 to the timing lags for the tongue tip and body gestures of non-contrastive dark [ɬ] reported by e.g. Sproat & Fujimura (1993) and Gick et al. (2006). Sproat & Fujimura (1993: 299, 303) report lags of 30–110ms between an earlier dorsal constriction and a later tongue tip constriction for coda [ɬ] in American English. Gick et al. (2006: Table 2) report smaller lags (≈ 20–25ms) for Western Canadian English, Quebec French,

[25]It should be acknowledged that our use of the symbols /iː uː ɔː/ to describe vowel quality obscures some variation in the quality of these vowels between speakers. Most notably, our Ulster speakers tend to front /uː ɔː/, often producing them as something like [ʌː ɛː]. Note, though, that the fronting of /uː ɔː/ should actually make palatalized /C^{j}/ seem *less* front, by comparison, in this context. Despite that fact, we still see clear fronting at C start relative to the preceding vowel for /C^{j}/ after /uː ɔː/ in Figure 16.

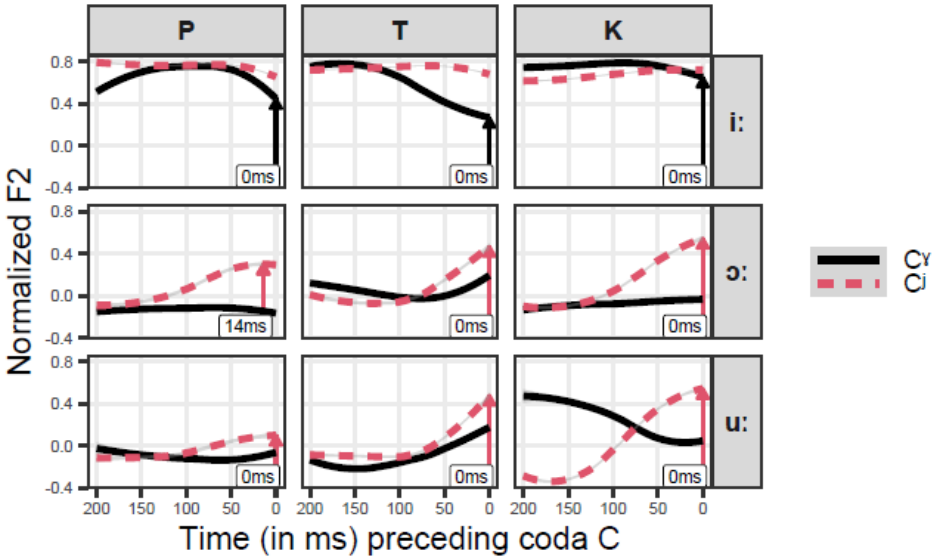


Figure 17: Pooled, speaker-normalized F2 trajectories across time, grouped by primary place of articulation and vowel context, for coda /VC#/. X-axis shows ms preceding closure of following coda consonant (up to 200ms). Arrows mark F2 extrema of loess curves for /iːCˠ uːCʲ ɔːCʲ/ (peak for /Cʲ/, trough for /Cˠ/).

and Squamish Salish.[26] Browman & Goldstein (1995) present 4 X-ray microbeam tokens of 'peel' [pʰiɬ] from one speaker, and find that tongue body retraction for [ɬ] peaks approximately 20–50ms before tongue tip closure for [ɬ] is completed, i.e. during the preceding vowel.

In contrast, Figure 17 shows that F2 extrema for these loess curves generally occur at the beginning of closure for the coda consonant, at least in those conditions where F2 transitions are most visible and interpretable due to a mismatch in backness between the vowel and following coda consonant. The only exception is /ɔːPʲ/, which has just a 14ms lag between the F2 peak during the vowel and the following coda /Pʲ/ in our data. There is clearly a contrast between these timing patterns, which imply (near) simultaneity between primary and secondary articulations in /VC/, and the more sequential timing reported for non-contrastive

[26]The difference in lag magnitudes reported by Sproat & Fujimura (1993) vs. Gick et al. (2006) likely reflects the fact that Sproat & Fujimura investigated /l/ darkening across a wider range of contexts, leading to greater variability in segment duration. Inter-speaker variation may also play a role, as Gick et al. (2006: 65) note. Gick et al. (2006) is also an ultrasound study with a relatively low framerate (30 fps ≈ 1 frame every 33ms), so their reported lag values are not particularly precise.

dark [ɬ] in Sproat & Fujimura (1993), and Gick et al. (2006) (see also Browman & Goldstein 1995; Krakow 1999).[27]

These patterns are incompatible with the claim that dorsal gestures for secondary /C^{ɣ} C^{j}/ articulations in coda position peak during the preceding vowel. Consequently, they are also incompatible with the claim that secondary /C^{ɣ} C^{j}/ articulations are timed to precede primary articulations in coda position. That is, the available evidence argues *against* the hypothesis that primary and secondary articulations are timed in a sequential, anti-phase relation in coda position.

The alternative hypothesis, which we adopt, is that secondary /C^{ɣ} C^{j}/ articulations are timed *in*-phase with either C start or C release, or with both, in coda position. (See also Shaw et al. 2021 on the timing of palatalization gestures with primary articulations in onset position.) We propose that this variability in timing reflects the fact that there is no single perceptually-optimal timing pattern in coda position, unlike onset position. In the absence of a unique, optimal timing pattern, speakers may be expected to show variation between different alternatives, giving rise to the gestural timing patterns we report here.

9 Discussion and conclusion

We began with the observation that secondary /C$^{(ɣ)}$ C^{j}/ contrasts are often eliminated in coda position and among labial consonants. This typological fact likely has a phonetic source in the relative difficulty of perceiving /C$^{(ɣ)}$ C^{j}/ contrasts in these contexts (e.g. Kochetov 2002; 2004; Ní Chiosáin & Padgett 2012; Padgett & Ní Chiosáin 2018), and the tendency to reduce articulatory gestures, including palatalization gestures, in codas (e.g. Kochetov 2002; 2006; 2009; Bennett et al. 2024).

Our results may extend this broad line of reasoning, to the extent that they suggest greater variability in timing strategies in coda position compared to onset position. Timing patterns which are more variable across contexts (such as place and syllable position) might be harder for listeners to attend to. If timing patterns also differ by speaker, this would be even more clearly true, because listeners cannot know ahead of time when the peak dorsal gestures for /C$^{(ɣ)}$ C^{j}/ contrasts will occur.

However, we have also argued that productions may be organized in ways that make life easier for listeners (whether this is intentional or not, see discussion

[27] Gick et al. (2006) similarly find simultaneous alignment of primary and secondary articulation in Serbo-Croatian /l^{ɣ}/, rather than anti-phase alignment as in English. They relate this finding to the fact of a palatalization contrast for laterals in the language (/l^{ɣ}/ vs. /l^{j}/).

in Section 3.1). In our data, secondary articulations in onset position are either timed to peak at C release, or appear to be held constant from C beginning to C release. In either case, gestural timing makes key acoustic information about /C$^{(ɣ)}$ C^{j}/ contrasts available at crucial perceptual landmarks, namely C release itself, and the following CV transition for onsets.

In coda position, consonants at different places of articulation show different timing patterns. The palatalization gesture for coda labial /P^{j}/ peaks at the VC transition. We have suggested that this reflects the perceptual importance of formant transitions for signaling /C$^{(ɣ)}$ C^{j}/ contrasts in labials, given the weak and relatively uninformative release bursts of labial stops. For coda coronal /T^{j}/, we instead find alignment of secondary articulations to C release; we have suggested that this reflects the acoustic and perceptual salience of coronal stop releases, which are very informative as to /C$^{(ɣ)}$ C^{j}/ contrast in Irish (e.g. Ní Chiosáin & Padgett 2012). If the secondary articulations for coda /T^{ɣ} K^{j} K^{ɣ}/ are held roughly constant until C release, as we've argued, that too may reflect the perceptual importance of release cues for coronal and dorsal stops.

We have seen that timing patterns for velarized /C^{ɣ}/ seem to be more ambiguous, or perhaps more varied, than the timing patterns found for palatalized /C^{j}/, particularly in onset position (as also found by Bennett et al. 2018). Why should this be? One idea pursued in Bennett et al. (2024) connects this finding to an independently known fact about palatal constrictions: they are more resistant to coarticulation (and more likely to induce coarticulation in other sounds) than many other kinds of sounds. Put differently, palatal constrictions exhibit a high degree of articulatory constraint (Recasens et al. 1997; Recasens 1999; Recasens & Espinosa 2009; Farnetani & Recasens 2010; Recasens & Rodríguez 2016). A reviewer suggests a second (possibly related) idea: that tongue body fronting may involve less complex muscle engagement than tongue body raising and backing (see e.g. Gick et al. 2012: 152–158). Finally, we know that lip rounding plays a role in enhancing velarization in Irish (Bennett et al. 2019). A fuller understanding of gestural timing in /C^{ɣ} C^{j}/ contrasts may therefore require attention to the movement of the lips, as well as the tongue. We leave all of these ideas for further research.

It should be borne in mind that our results here pertain only to stop consonants. As discussed in Section 3.3, previous findings on Russian suggest that timing facts may depend on a consonant's manner of articulation.[28] Unlike stops, other consonants such as fricatives, nasals and liquids do not have release bursts, and

[28]Bennett et al. (2018) do not find a difference in timing between Irish stops and fricatives, but they only examine consonants in word-/syllable-initial position.

they may carry internal cues to the palatalization contrast. Whether a lateral is palatalized or velarized, for example, is easily audible during closure. Hence, we do not readily expect our findings to extend to other manners of articulation and further research on this topic is needed.

Our characterization of the issues has been oversimplified in an important way: we have discussed cues to palatalization without considering the larger phonetic context in which a word occurs. If a word like *píob* /pʲiːbˠ/ occurs in isolation, then indeed the cues to palatalization of the initial consonant lie only at the consonant's release, and those of the final consonant lie at both the VC transition and the consonant's release (assuming the consonant is released, as it generally would be in Irish). However, our target words were in a different context, since we used the carrier sentence [ˈdˠuːrʲtʲ ˈiːfˠə ___ əˈnˠʊrˠə] 'Aoife said ___ last year'. They were preceded by a vowel-final word and followed by a vowel-initial one. In such a context, a listener could, in principle, have access to VC formant transitions, a burst, and CV formant transitions for *both* initial and final consonants. Why, then, did we not find identical patterns of timing for initial and final consonants?

Had we manipulated the phrasal context, independently varying the immediate segmental context of initial versus final consonants, and had this had effects on timing, we might have concluded that speakers can adjust the relative timing of primary and secondary articulations 'online'; that is, speaker adjustments can be made in a way that adapts to phrasal context. We did not manipulate phrasal context. However, since we in fact found *different* patterns of timing for initial and final consonants, even though the segmental context was the same (V_V), our conclusion might rather be that speakers *cannot* adjust the relative timing of articulations in this way. Though somewhat speculative, this conclusion seems well in line with the hypothesis that word-internal gestural timing is lexically-specified, and so less variable than the relative timing of gestures across a word boundary (Browman & Goldstein 1988; Byrd 1996; Browman & Goldstein 2000), an idea supported by the results of Cho (2001) (see Strycharczuk 2019; Mousikou et al. 2021 and references therein for further evidence and discussion). The question then becomes: what patterns of timing, fixed in the lexicon, would be most perceptually adaptive? Such patterns of timing would have to facilitate the perception of cues to palatalization not only in the most accommodating of environments (such as that of our carrier sentence); they would have to succeed in *all* environments, including ones where a consonant is preceded or followed by silence or another consonant (see Flemming 2001: 30–31 for a similar point). It is possible that lexically-specified gestural timing 'gravitates' to settings that work best for listeners across all environments.

Overall, our data suggests that relatively low-level timing patterns in the production of Irish stop consonants may be structured in such a way as to maximize the perceptual salience and recoverability of cues to /$C^{(\gamma)}$ C^{j}/ contrasts.

Abbreviations

SG.	singular
PL.	plural
GEN.	genitive
fps	frames per second
P	labial stop
T	coronal stop
K	dorsal stop

Acknowledgements

Thanks to audiences at HISPhonCog and ICPhS in 2023, as well as Washington University in St. Louis and the University of Connecticut, for feedback on earlier versions of this work. We are also grateful to Nick Van Handel, and a number of undergraduate research assistants, for helping us process the data presented on here. This paper was improved by helpful comments from two anonymous reviewers, who we thank as well.

Contributions

Ryan Bennett and Jaye Padgett contributed to conceptualization of the project, design of the methodology and items, data collection, data analysis, and writing. Grant McGuire and Máire Ní Chiosáin contributed to conceptualization of the project, design of the methodology and items, data collection, reviewing of the manuscript, and editing. Jennifer Bellik contributed to data analysis.

Appendix A Words used in the study

Table 9: Onset/word-initial position

	p/b	t/d	k
iː	bˠiː ‘yellow’ pʲiːsˠə ‘piece’	tˠiː ‘straw’ dʲiːnˠ ‘roof’	kˠiː ‘way’ kʲiːrˠ ‘a comb’
uː	pˠuːkˠə ‘ghost’ bʲuː ‘would be worth’	tˠuːsˠ ‘a start’ tʲuːsˠ ‘thickness’	kˠuː ‘a hound’ kʲuːnˠəsˠ ‘quiet (noun)’
ɔː	pˠɔː ‘pay, wages’ pʲɔːnˠ ‘pen’	tˠɔː ‘is (be)’ tʲɔːnˠ ‘tight’	kˠɔːrˠ ‘car’ kʲɔːnˠ ‘one/head’

Table 10: Coda/word-final position

	p/b	t/d	k/g
iː	pʲiːbˠ ‘pipes (GEN.PL)’ pʲiːbʲ ‘pipes’	iːədˠ ‘them’ trʲiːdʲ ‘through’	iːəkˠ ‘pay’ bʲeːkʲ‘shout’
uː	lˠuːbˠ ‘stitch’ lˠuːbʲ ‘stitch (DAT.SG./variant)’	uːdˠ ‘that’ klˠuːdʲ ‘covering’	glʲuːkˠ ‘peep’ bˠuːkʲ ‘pinnacle’
ɔː	lˠɔːbˠ ‘mud (variant)’ lˠɔːbʲ ‘mud’	stˠɔːtˠ ‘state’ ɔːtʲ ‘place’	fˠɔːgˠ ‘leave’ rˠɔːgʲ ‘sudden rush’

Appendix B Token counts

Table 11: Token counts by speaker, place, and secondary articulation

Dialect	Speaker	C place	/Cˠ Cʲ/	*n*
Connacht	1	Labial	Cˠ	89
Connacht	1	Labial	Cʲ	90
Connacht	2	Labial	Cˠ	72
Connacht	2	Labial	Cʲ	72

Dialect	Speaker	C place	/C$^{\gamma}$ C^{j}/	*n*
Connacht	2	Coronal	C$^{\gamma}$	71
Connacht	2	Coronal	C^{j}	66
Connacht	2	Dorsal	C$^{\gamma}$	72
Connacht	2	Dorsal	C^{j}	71
Connacht	3	Labial	C$^{\gamma}$	90
Connacht	3	Labial	C^{j}	89
Connacht	3	Dorsal	C$^{\gamma}$	84
Connacht	3	Dorsal	C^{j}	87
Munster	1	Labial	C$^{\gamma}$	90
Munster	1	Labial	C^{j}	90
Munster	1	Coronal	C$^{\gamma}$	85
Munster	1	Coronal	C^{j}	83
Munster	2	Labial	C$^{\gamma}$	86
Munster	2	Labial	C^{j}	88
Munster	2	Coronal	C$^{\gamma}$	79
Munster	2	Coronal	C^{j}	85
Ulster	1	Labial	C$^{\gamma}$	90
Ulster	1	Labial	C^{j}	90
Ulster	1	Coronal	C$^{\gamma}$	90
Ulster	1	Coronal	C^{j}	86
Ulster	2	Labial	C$^{\gamma}$	90
Ulster	2	Labial	C^{j}	90
Ulster	2	Coronal	C$^{\gamma}$	90
Ulster	2	Coronal	C^{j}	84
Ulster	2	Dorsal	C$^{\gamma}$	90
Ulster	2	Dorsal	C^{j}	90

References

Baayen, R. Harald, Douglas Davidson & Douglas Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). 390–412. DOI: 10.1016/j.jml.2007.12.005.

Barr, Dale, Roger Levy, Christoph Scheepers & Harry Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278. DOI: 10.1016/j.jml.2012.11.001.

Barreda, Santiago. 2021. Fast Track: Fast (nearly) automatic formant-tracking using Praat. *Linguistics Vanguard* 7(1). 20200051. DOI: 10.1515/lingvan-2020-0051.

Barreda, Santiago & Terrance Nearey. 2018. A regression approach to vowel normalization for missing and unbalanced data. *The Journal of the Acoustical Society of America* 144(1). 500–520. DOI: 10.1121/1.5047742.

Bateman, Nicoleta. 2011. On the typology of palatalization. *Language and Linguistics Compass* 5(8). 588–602. DOI: 10.1111/j.1749-818x.2011.00294.x.

Bennett, Ryan, Máire Ní Chiosáin, Jaye Padgett & Grant McGuire. 2018. An ultrasound study of Connemara Irish palatalization and velarization. *Journal of the International Phonetic Association* 48(3). 261–304. DOI: 10.1017/S0025100317000494.

Bennett, Ryan, Jaye Padgett, Máire Ní Chiosáin, Grant McGuire & Jennifer Bellik. 2019. Contrast enhancement and cue trading in Irish secondary consonant articulations. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, 3584–3552. Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Bennett, Ryan, Jaye Padgett, Máire Ní Chiosáin, Grant McGuire & Jennifer Bellik. 2024. Effects of syllable position and place of articulation on secondary dorsal contrasts: An ultrasound study of Irish. *Journal of Phonetics* 107. 101368. DOI: 10.1016/j.wocn.2024.101368.

Bird, Sonya & Natallia Litvin. 2021. Belarusian. *Journal of the International Phonetic Association* 51(3). 450–467. DOI: 10.1017/s0025100319000288.

Biteeva, Ekaterina. 2021. *Complexité et contrôle du geste linguo-palatal sous l'éclairage de sa variabilité. le cas de la palatalisation en russe. aspects phonétiques et phonologiques.* Université Grenoble Alpes. (Doctoral dissertation).

Boersma, Paul & David Weenink. 2020. *Praat: doing phonetics by computer (version 6.1.05)*. Computer program. http://www.praat.org/ (5 June, 2025).

Bolker, Benjamin, Mollie Brooks, Connie Clark, Shane Geange, John Poulsen, M. Henry Stevens & Jada-Simone White. 2009. Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in ecology and evolution* 24(3). 127–135. DOI: 10.1016/j.tree.2008.10.008.

Browman, Catherine & Louis Goldstein. 1988. Some notes on syllable structure in Articulatory Phonology. *Phonetica* 45(2-4). 140–155. DOI: 10.1159/000261823.

Browman, Catherine & Louis Goldstein. 1995. Gestural syllable position effects in American English. In Fredericka Bell-Berti & Lawrence J. Raphael (eds.), *Producing speech: Contemporary issues (for K. S. Harris)*, 19–33. Woodbury: AIP Press.

Browman, Catherine & Louis Goldstein. 2000. Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP. Bulletin de la communication parlée* 5. 25–34. DOI: 10.1515/tlr.2007.007.

Byrd, Dani. 1996. A phase window framework for articulatory timing. *Phonology* 13(2). 139–169. DOI: 10.1017/S0952675700002086.

Byrd, Dani, Stephen Tobin, Erik Bresch & Shrikanth Narayanan. 2009. Timing effects of syllable structure and stress on nasals: A real-time MRI examination. *Journal of Phonetics* 37(1). 97–110. DOI: 10.1016/j.wocn.2008.10.002.

Chernick, Michael. 2011. *Bootstrap methods: A guide for practitioners and researchers*. Hoboken: John Wiley & Sons.

Chitoran, Ioana. 2013. *The phonology of Romanian: A constraint-based approach*. Berlin: de Gruyter.

Cho, Taehong. 2001. Effects of morpheme boundaries on intergestural timing: Evidence from Korean. *Phonetica* 58(3). 129–162. DOI: 10.1159/000056196.

Clayton, Ian. 2010. *On the natural history of preaspirated stops*. University of North Carolina at Chapel Hill. (Doctoral dissertation).

Czaplicki, Bartłomiej. 2010. Palatalized labials in Polish dialects: An evolutionary perspective. *Journal of Slavic linguistics* 18(2). 259–290. DOI: 10.1353/jsl.2010.0002.

De Bhaldraithe, Tomás. 1975. *The Irish of Cois Fhairrge, Co. Galway: A phonetic study*. First published 1945. Dublin: Dublin Institute for Advanced Studies.

De Búrca, Seán. 1977. Syllabicity and palatalization. *Studia Celtica* 12. 396–404.

Eska, Joseph F. 2017. The dialectology of Celtic. In Jared Klein, Brian Joseph, Matthias Fritz & Mark Wenthe (eds.), *Handbook of comparative and historical Indo-European linguistics*, 1264–1274. Berlin: de Gruyter Mouton. DOI: 10.1515/9783110523874-027.

Farnetani, Edda & Daniel Recasens. 2010. Coarticulation and connected speech processes. In William Hardcastle, John Laver & Fiona Gibbon (eds.), *The handbook of phonetic sciences*, 316–352. Oxford: Blackwell. DOI: 10.1002/9781444317251.ch9.

Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18(1). 7–44. DOI: 10.1017/S0952675701004006.

Frick, Robert. 1995. Accepting the null hypothesis. *Memory & Cognition* 23(1). 132–138. DOI: 10.3758/bf03210562.

Fujimura, Osamu, Marian Macchi & Lynn Streeter. 1978. Perception of stop consonants with conflicting transitional cues. *Language and Speech* 21. 337–343. DOI: 10.1177/002383097802100408.

Gelman, Andrew & Jennifer Hill. 2006. *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Gick, Bryan, Fiona Campbell, Sunyoung Oh & Linda Tamburri-Watt. 2006. Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of phonetics* 34(1). 49–72. DOI: 10.1016/j.wocn.2005.03.005.

Gick, Bryan, Ian Wilson & Donald Derrick. 2012. *Articulatory phonetics*. Malden: Wiley-Blackwell.

Gick, Bryan, Ian Wilson, Karsten Koch & Clare Cook. 2004. Language-specific articulatory settings: Evidence from inter-utterance rest position. *Phonetica* 61(4). 220–233. DOI: 10.1159/000084159.

Goldstein, Louis, Dani Byrd & Elliot Saltzman. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In Michael A. Arbib (ed.), *Action to language via the mirror neuron system*, 215–249. Cambridge: Cambridge University Press. DOI: 10.1017/cbo9780511541599.008.

Green, Antony Dubach. 2001. The prosodic representation of clitics in Irish. In Birgit Gerlach & Janet Grijzenhout (eds.), *Clitics in phonology, morphology and syntax*, 181–218. Amsterdam: John Benjamins. DOI: 10.1075/la.36.09dub.

Greene, David. 1973. The growth of palatalization in Old Irish. *Transactions of the Philological Society* 72(1). 127–136. DOI: 10.1111/j.1467-968x.1973.tb01017.x.

Harrell Jr., Frank E. 2023. *Hmisc: Harrell Miscellaneous*. R package version 5.0-1. https://CRAN.R-project.org/package=Hmisc (5 June, 2025).

Hickey, Raymond. 2011. *The dialects of Irish: study of a changing landscape*. Berlin: Walter de Gruyter.

Hueber, Thomas, Gerard Chollet, Bruce Denby & Maureen Stone. 2008. Acquisition of ultrasound, video and acoustic speech data for a silent-speech interface application. In Rudolph Sock, Susanne Fuchs & Yves Laprie (eds.), *Proceedings of International Seminar on Speech Production (ISSP) (strasbourg)*, vol. 8, 365–369. INRIA.

International Phonetic Association (ed.). 1999. *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.

Iskarous, Khalil & Darya Kavitskaya. 2010. The interaction between contrast, prosody, and coarticulation in structuring phonetic variability. *Journal of Phonetics* 38(4). 625–639. DOI: 10.1016/j.wocn.2010.09.004.

Iskarous, Khalil & Darya Kavitskaya. 2018. Sound change and the structure of synchronic variability: Phonetic and phonological factors in Slavic palatalization. *Language* 94(1). 43–83. DOI: 10.1353/lan.0.0226.

Jackson, Kenneth Hurlstone. 1967. Palatalization of labials in the Gaelic languages. In Wolfgang Meid (ed.), *Beiträge zur Indogermanistik und Keltologie*, 179–192. Innsbruck: Sprachwissenschaftliches Institut der Universität Innsbruck.

Katsika, Argyro. 2016. The role of prominence in determining the scope of boundary-related lengthening in Greek. *Journal of Phonetics* 55. 149–181. DOI: 10.1016/j.wocn.2015.12.003.

Kingston, John. 1990. Articulatory binding. In John Kingston & Mary Beckman (eds.), *Papers in laboratory phonology I: Between the grammar and the physics of speech*, 406–434. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511627736.023.

Kingston, John. 2019. The interface between phonetics and phonology. In William F. Katz & Peter F. Assmann (eds.), *The Routledge handbook of phonetics*, 359–400. New York: Routledge. DOI: 10.4324/9780429056253-14.

Kingston, John & Randy Diehl. 1994. Phonetic knowledge. *Language* 70(3). 419–454. DOI: 10.2307/416481.

Kirkham, Sam & Claire Nance. 2022. Diachronic phonological asymmetries and the variable stability of synchronic contrast. *Journal of Phonetics* 94. 101176. DOI: 10.1016/j.wocn.2022.101176.

Kochetov, Alexei. 2002. *Production, perception and emergent phonotactic patterns: A case of contrastive palatalization*. London: Routledge.

Kochetov, Alexei. 2004. Perception of place and secondary articulation contrasts in different syllable positions: Language-particular and language-independent asymmetries. *Language and Speech* 47(4). 351–382. DOI: 10.1177/00238309040470040201.

Kochetov, Alexei. 2005. Phonetic sources of phonological asymmetries: Russian laterals and rhotics. In Claire Gurski (ed.), *Actes du congrès de l'ACL 2005*.

Kochetov, Alexei. 2006. Syllable position effects and gestural organization: Evidence from Russian. In Louis Goldstein, D.H. Whalen & Catherine Best (eds.), *Papers in laboratory phonology VIII*, 565–588. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110197211.3.565.

Kochetov, Alexei. 2009. Phonetic variation and gestural specification: Production of Russian consonants. In Aditi Lahiri, Frank Kügler, Caroline Féry & Ruben van de Vijver (eds.), *Variation and gradience in phonetics and phonology*, 43–70. Berlin: Walter de Gruyter. DOI: 10.1515/9783110219326.43.

Krakow, Rena A. 1999. Physiological organization of syllables: A review. *Journal of Phonetics* 27(1). 23–54. DOI: 10.1006/jpho.1999.0089.

Kruschke, John. 2013. Bayesian estimation supersedes the t test. *Journal of Experimental Psychology* 142(2). 573–603. DOI: 10.1037/e502412013-055.

Kuznetsova, Alexandra, Per Bruun Brockhoff, Rune Haubo Bojesen Christensen & Sofie Pødenphant Jensen. 2020. *lmerTest: Tests in linear mixed effects models*. Version 3.1-3, retrieved from http://CRAN.R-project.org/package=lmerTest. DOI: 10.32614/cran.package.lmertest.

Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Malden: Blackwell.

Lee-Kim, Sang-Im, Lisa Davidson & Sangjin Hwang. 2013. Morphological effects on the darkness of English intervocalic /l/. *Laboratory Phonology* 4(2). 475–511. DOI: 10.1515/lp-2013-0015.

Lewin, Christopher. 2019. *Aspects of the historical phonology of Manx*. University of Edinburgh. (Doctoral dissertation).

Li, Min, Chandra Kambhamettu & Maureen Stone. 2005. Automatic contour tracking in ultrasound images. *Clinical Linguistics & Phonetics* 19(6-7). 545–554. DOI: 10.1080/02699200500113616.

Liljencrants, Johan & Björn Lindblom. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48(4). 839–862. DOI: 10.2307/411991.

Lindblom, Björn. 1986. Phonetic universals in vowel systems. In John J. Ohala & Jeri Jaeger (eds.), *Experimental phonology*, 13–44. Orlando: Academic Press.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H theory. In William Hardcastle & Alain Marchal (eds.), *Speech production and speech modelling*, 403–439. Dordrecht: Kluwer. DOI: 10 . 1007 / 978 - 94 - 009 - 2037-8_16.

Lüdecke, Daniel, Mattan S. Ben-Shachar, Indrajeet Patil, Philip Waggoner & Dominique Makowski. 2021. performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software* 6(60). 3139. DOI: 10.21105/joss.03139.

Makowski, Dominique, Mattan S. Ben-Shachar & Daniel Lüdecke. 2019. bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software* 4(40). 1541. DOI: 10.21105/joss.01541.

Malmi, Anton, Pärtel Lippus & Einar Meister. 2023. Spectral and temporal properties of Estonian palatalization. *Journal of the International Phonetic Association* 53(3). 748–773. DOI: 0.1017/S0025100321000360.

Martinet, André. 1952. Function, structure, and sound change. *Word* 8(1). 1–32. DOI: 10.1080/00437956.1952.11659416.

Martinet, André. 1964. *Économie des changements phonétiques: Traité de phonologie diachronique*. Berne: A. Francke.

Matuschek, Hannes, Reinhold Kliegl, Shravan Vasishth, R. Harald Baayen & Douglas Bates. 2017. Balancing Type I error and power in linear mixed models. *Journal of Memory and Language* 94. 305–315. DOI: 10.1016/j.jml.2017.01.001.

McCone, Kim. 1996. *Towards a relative chronology of ancient and medieval Celtic sound change*. Maynooth: The Cardinal Press.

Monahan, Philip & William Idsardi. 2010. Auditory sensitivity to formant ratios: Toward an account of vowel normalisation. *Language and Cognitive Processes* 25(6). 808–839. DOI: 10.1080/01690965.2010.490047.

Morey, Richard D. & Jeffrey N. Rouder. 2023. *BayesFactor: Computation of Bayes Factors for Common Designs*. R package version 0.9.12-4.5. https://CRAN.R-project.org/package=BayesFactor (5 June, 2025).

Mousikou, Petroula, Patrycja Strycharczuk, Alice Turk & James Scobbie. 2021. Coarticulation across morpheme boundaries: An ultrasound study of past-tense inflection in Scottish English. *Journal of Phonetics* 88. 101101. DOI: 10.1016/j.wocn.2021.101101.

Nam, Hosung, Louis Goldstein & Elliot Saltzman. 2009. Self-organization of syllable structure: A coupled oscillator model. In François Pellegrino, Egidio Marsico, Ioana Chitoran & Christophe Coupé (eds.), *Approaches to phonological complexity*, 297–328. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110223958.297.

Nance, Claire & Roibeard Ó Maolalaigh. 2021. Scottish Gaelic. *Journal of the International Phonetic Association* 51(2). 261–275. DOI: 10.1017/s002510031900015x.

Ní Chasaide, Ailbhe. 1995. Irish. *Journal of the International Phonetic Association* 25(1). 34–39. DOI: 10.1017/S0025100300000189.

Ní Chasaide, Ailbhe & Geraldine Fealy. 1991. A preliminary investigation of palatal/palatalised consonants of Irish. Unpublished ms.

Ní Chiosáin, Máire. 1991. *Topics in the phonology of Irish*. University of Massachusetts, Amherst. (Doctoral dissertation).

Ní Chiosáin, Máire. 1994. Irish palatalisation and the representation of place features. *Phonology* 11(1). 89–106. DOI: 10.1017/s0952675700001858.

Ní Chiosáin, Máire. 1999. Syllables and phonotactics in Irish. In Harry van der Hulst & Nancy Ritter (eds.), *The syllable: Views and facts*, 551–575. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110806793.551.

Ní Chiosáin, Máire. Forthcoming. Modern Irish phonology and phonetics. In Joseph Eska, Silva Nurmio, Peadar Ó Muircheartaigh & Paul Russell (eds.), *Palgrave handbook of Celtic languages and linguistics*. Palgrave Macmillan.

Ní Chiosáin, Máire & Jaye Padgett. 2012. An acoustic and perceptual study of Connemara Irish palatalization. *Journal of the International Phonetic Association* 42(2). 171–191. DOI: 10.1017/s0025100312000059.

Ó Siadhail, Mícheál. 1991. *Modern Irish: grammatical structure and dialectal variation*. Cambridge: Cambridge University Press.

Oberpriller, Johannes, Melina de Souza Leite & Maximilian Pichler. 2022. Fixed or random? On the reliability of mixed-effects models for a small number of levels in grouping variables. *Ecology and Evolution* 12(7). e9062. DOI: 10.1002/ece3.9062.

Ohala, John J. 1981. The listener as a source of sound change. In Carrie S. Masek, Roberta A. Hendrick & Mary Frances Miller (eds.), *Papers from the parasession on language and behavior*, 178–203. Chicago: Chicago Linguistics Society. DOI: 10.1075/cilt.323.05oha.

Ohala, John J. 1990. The phonetics and phonology of aspects of assimilation. In John Kingston & Mary Beckman (eds.), *Papers in laboratory phonology I: Between the grammar and the physics of speech*, 258–275. Cambridge: Cambridge University Press. DOI: 10.1017/cbo9780511627736.014.

Ohala, John J. 1993a. Coarticulation and phonology. *Language and Speech* 36(2-3). 155–170. DOI: 10.1177/002383099303600303.

Ohala, John J. 1993b. The phonetics of sound change. In Charles Jones (ed.), *Historical linguistics: Problems and perspectives*, 237–278. London: Longman. DOI: 10.4324/9781315845814.

Operstein, Natalie. 2010. Consonant structure and prevocalization. *Consonant Structure and Prevocalization*. DOI: 10.1075/cilt.312.

Padgett, Jaye, Ryan Bennett, Grant McGuire, Máire Ní Chiosáin & Jennifer Bellik. 2023. The timing of secondary dorsal articulations across syllable positions in Irish. In Radek Skarnitzl & Jan Volín (eds.), *Proceedings of the 20th International Congress of Phonetic Sciences, Prague, Czech Republic 2023*, 2209–2213. Prague: Guarant International.

Padgett, Jaye & Scott Myers. 2014. Domain generalization in artificial language learning. *Phonology* 31(3). 399–433. DOI: 10.1017/s0952675714000207.

Padgett, Jaye & Máire Ní Chiosáin. 2018. The perception of a secondary palatalization contrast: A preliminary comparison of Russian and Irish. In Ryan Bennett, Adrian Brasoveanu, Dhyana Buckley, Nick Kalivoda, Shigeto Kawahara, Grant McGuire & Jaye Padgett (eds.), *A festschrift for Junko Itô and Armin Mester*.

Santa Cruz: Department of Linguistics, University of California, Santa Cruz. https://itomestercelebration.sites.ucsc.edu/ (5 June, 2025).

Parrell, Benjamin. 2012. The role of gestural phasing in Western Andalusian Spanish aspiration. *Journal of Phonetics* 40(1). 37–45. DOI: 10.1016/j.wocn.2011.08.004.

Pinheiro, José & Douglas Bates. 2000. *Mixed-effects models in S and S-PLUS*. New York: Springer.

Proctor, Michael. 2009. *Gestural characterization of a phonological class: The liquids*. Yale. (Doctoral dissertation).

Purcell, Edward T. 1979. Formant frequency patterns in Russian VCV utterances. *Journal of the Acoustical Society of America* 66(6). 1691–1702. DOI: 10.1121/1.383641.

R Development Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Version 4.0.2. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org/ (5 June, 2025).

Recasens, Daniel. 1999. Lingual coarticulation. In William Hardcastle & Nigel Hewlett (eds.), *Coarticulation: Theory, data, and techniques*, 80–104. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511486395.005.

Recasens, Daniel & Aina Espinosa. 2009. An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *The Journal of the Acoustical Society of America* 125(4). 2288–2298. DOI: 10.1121/1.3089222.

Recasens, Daniel, Maria Dolors Pallarès & Jordi Fontdevila. 1997. A model of lingual coarticulation based on articulatory constraints. *The Journal of the Acoustical Society of America* 102(1). 544–561. DOI: 10.1121/1.419727.

Recasens, Daniel & Clara Rodríguez. 2016. A study on coarticulatory resistance and aggressiveness for front lingual consonants and vowels using ultrasound. *Journal of Phonetics* 59. 58–75. DOI: 10.1016/j.wocn.2016.09.002.

Scatton, Ernest. 1993. Bulgarian. In Greville Corbett & Bernard Comrie (eds.), *The Slavonic languages*, 188–248. London: Routledge.

Scobbie, James, Alan Wrench & Marietta van der Linden. 2008. Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement. In Rudolph Sock, Susanne Fuchs & Yves Laprie (eds.), *Proceedings of International Seminar on Speech Production (ISSP) (strasbourg)*, vol. 8, 373–376. INRIA.

Shaw, Jason, Sejin Oh, Karthik Durvasula & Alexei Kochetov. 2021. Articulatory coordination distinguishes complex segments from segment sequences. *Phonology* 38(3). 437–477. DOI: 10.1017/s0952675721000269.

Short, David. 1993. Czech. In Greville Corbett & Bernard Comrie (eds.), *The Slavonic languages*, 455–532. London: Routledge.

Silverman, Daniel. 1997. Laryngeal complexity in Otomanguean vowels. *Phonology* 14(2). 235–261. DOI: 10.1017/S0952675797003412.

Sproat, Richard & Osamu Fujimura. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21. 291–311. DOI: 10.1016/s0095-4470(19)31340-3.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. In Keith Johnson & Elizabeth Hume (eds.), *The role of speech perception in phonology*, 219–250. New York: Academic Press. DOI: 10.1163/9789004454095_013.

Stevens, Kenneth N. & Dennis Klatt. 1974. Role of formant transitions in the voiced-voiceless distinction for stops. *The Journal of the Acoustical Society of America* 55(3). 653–659. DOI: 10.1121/1.1914578.

Stifter, David. 2017. The phonology of Celtic. In Jared Klein, Brian Joseph, Matthias Fritz & Mark Wenthe (eds.), *Handbook of comparative and historical Indo-European linguistics*, 1188–1202. Berlin: de Gruyter Mouton. DOI: 10.1515/9783110523874-023.

Stoll, Taja. 2017. *Articulatory analysis of palatalised rhotics in Russian: Implications for sound change*. Ludwig-Maximilians-Universität, München. (Doctoral dissertation).

Strycharczuk, Patrycja. 2019. Phonetic detail and phonetic gradience in morphological processes. In *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press. DOI: 10.1093/acrefore/9780199384655.013.616.

Sung, Jae-Hyun, Diana Archangeli, Sam Johnston, Ian Clayton, Muriel Fisher, Michael Hammond & Andrew Carnie. 2018. Individual variation in lexical palatalization: Articulatory evidence from Scottish Gaelic. *Eoneohag* 80. 23–51. DOI: 10.17290/jlsk.2018..80.23.

Takatori, Yuki. 1997. *A study of constraint interaction in Slavic phonology*. New Haven: Yale University.

Ternes, Elmar. 2006. *The phonemic analysis of Scottish Gaelic (3rd ed.)* Dublin: Dublin Institute for Advanced Studies.

Tomaschek, Fabian, Peter Hendrix & R. Harald Baayen. 2018. Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics* 71. 249–267. DOI: 10.1016/j.wocn.2018.09.004.

Turk, Alice, Satsuki Nakai & Mariko Sugahara. 2006. Acoustic segment durations in prosodic research: A practical guide. In Stefan Sudhoff, Denisa Lenertova, Roland Meyer, Sandra Pappert, Petra Augurzky, Ina Mleinek, Nicole Richter &

Johannes Schließer (eds.), *Methods in empirical prosody research*, 1–28. Berlin: de Gruyter. DOI: 10.1515/9783110914641.1.

Turton, Danielle. 2017. Categorical or gradient? An ultrasound investigation of /l/-darkening and vocalization in varieties of English. *Laboratory Phonology* 8(1). 13. DOI: 10.5334/labphon.35.

Wrench, Alan. 2008. *Ultrasound Stabilisation Headset users manual*. Version 1.5. Articulate Instruments Ltd. http : / / materials . articulateinstruments . com / Manuals/ (5 June, 2025).

Wright, Richard. 1996. *Consonant clusters and cue preservation in Tsou*. UCLA. (Doctoral dissertation).

Wright, Richard. 2004. A review of perceptual cues and cue robustness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically based phonology*, 34–57. Cambridge: Cambridge University Press. DOI: 10 . 1017 / cbo9780511486401.002.

Part III

Language change, historical linguistics, and grammaticalization

Chapter 9

Comparing the syntactic complexity of Gaeltacht and urban Irish-Language broadcasters

Brian Ó Broin
William Paterson University, NJ

Irish, while nominally a national language, is primarily spoken as a native language in rural enclaves called Gaeltachtaí. The language has recently experienced an urban revival, however, with Irish now being spoken as a domestic language in towns and cities all over Ireland. Following on my previous work in phonetics and morphology, this paper analyzes the two varieties of Irish (rural and urban) with particular emphasis on syntax and lexicon, with a view to discovering whether they are distinguishable using those features. I undertake a syntactic and lexical analysis of Gaeltacht and Urban speakers of Modern Irish. In the case of syntax, I am applying the well-known D-Level complexity scale of Rosenberg and Abbeduto to a number of transcriptions from speakers on urban and Gaeltacht radio stations (and adapting the scale for Irish syntax), while in the case of lexicon I perform a simple lexical diversity analysis on the same transcriptions using TTR (Type-Token Ratio) calculation. A complexity scale for Irish has been a major desideratum for several decades, and this article intends to provide such a scale. Similarly, no TTR analysis of Irish adults has yet been done, and this analysis will be a first.

1 Introduction

Rosenberg & Abbeduto (1987) developed a complexity scale for English sentences to analyze the language of adults with intellectual disabilities. Scholars quickly realized, however, that this complexity scale could be used for the linguistic analysis of any speakers, and the scale has been retained, with adaptations and revisions, as a useful yardstick for gauging developmental level. This paper sets

Brian Ó Broin. 2025. Comparing the syntactic complexity of Gaeltacht and urban Irish-Language broadcasters. In Andrew Carnie, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond & Luis Irizarry (eds.), *Foundational approaches to Celtic linguistics*, 345–368. Berlin: Language Science Press. DOI: 10.5281/zenodo.15654873

out to apply an adapted version of the Rosenberg and Abbeduto scale to urban and Gaeltacht broadcasters of Irish with the goal of determining if there are measurable differences between the syntactic complexity of broadcasters in the two communities.

There are two principal Irish language communities. The traditionally recognized group is that of the "Gaeltacht," a government-recognized area where the number of Irish speakers exceeds a certain threshold.[1] A second, larger, non-Gaeltacht community consists of mostly urban language activists and learners who frequently equate their language with identity in much the same way the Basque and Catalan communities do in France and Spain. This community, existing through multiple generations since even before the foundation of the Irish state (Edwards 2016),[2] has been much supported in recent decades by the growth of Irish-medium education under the Gaelscoileanna ("Gaelic Schools") movement, which has now coalesced with secondary-level "Gaelcholáistí" ("Gaelic Colleges") under the "Gaeloideachas" ("Gaelic Education") banner. Somewhere in the region of 43,000 pupils attend such Irish-medium schools outside the Gaeltacht.[3]

Neither community is monolithic. Some officially recognised Gaeltacht areas like Baile Cláir in Co. Galway now have certainly fewer than 5% native speakers (Ó Cinnseala 2020), while vigorous, but unrecognised, "nua-Ghaeltacht" ("neo-Gaeltacht") areas like Carn Tóchair in Co. Derry may well have more.[4] The Gaeltacht community represents an unbroken language tradition that dates back scores of generations to the arrival of Celtic culture in Ireland. Its speakers, speaking what Feargal Ó Béarra termed "Traditional Late Modern Irish" or "Non-Traditional Late Modern Irish" (Ó Béarra 2007), mostly take their language for granted and have little or no interest in language preservation or promotion. Gaeltacht activism exists, of course, and is the predominant reason for the existence of the national radio and television broadcasters RTÉ Raidió na Gaeltachta and TG4 (Watson 2002: 742–745), but most native speakers in the Gaeltacht speak Irish

[1] Definitions and thresholds have changed and fluctuated since the foundation of the Irish state, but the constant is that these areas receive more government funding than other areas for Irish-language education and for projects that are seen to preserve the language community. See Ó Gadhra (1988) for a good introduction to the subject.

[2] Because most members of this community raise their children in English, however, the community's members come and go from generation to generation without transferring their Irish to the next generation.

[3] Another 14,000 attend school in the Gaeltacht itself, where Irish is the default language of schools (Gaeloideachas 2023).

[4] See Armstrong (2012). Localized language statistics for United Kingdom districts are unavailable.

simply because, as a mother tongue and community language for them, it is easier for them than English. The community has been much supported by the existence of the afore-mentioned Gaeltacht-focused radio station, which broadcasts almost entirely from its three studios in the Gaeltacht areas of Kerry, Galway, and Donegal. Although the station is available nationally and internationally on RTÉ's broadcast network,[5] the six officially-recognized Gaeltacht areas around the country are considered its primary audience.[6]

The non-Gaeltacht language community is more amorphous, and though this author refers to the community as urban, there are of course non-Gaeltacht speakers who live in rural communities. That said, this study mostly confines itself to the non-Gaeltacht urban communities of Dublin and Belfast, cities where enough language activists and daily speakers exist to justify the existence of print and broadcast media to service them.[7] These speakers vary considerably in ability and interest, from native speakers raised through Irish in the city and educated in Irish-medium schools to adult learners who may never have learned a word until enrolling in an evening class. Both urban communities are serviced by local radio stations broadcasting around the clock: Raidió Fáilte in Belfast and Raidió na Life in Dublin. Dublin, Belfast, and Cork speakers are often easy to identify by their dialects.[8] Some might argue that this is merely the effect of strong English accents bleeding over into the second language, but these speakers are nevertheless recognizable.

The urban community of Irish speakers is growing, with language-centered events taking place very regularly in many towns and cities,[9] and there has been sociolinguistic analysis of those who participate in such events. Walsh & O'Rourke (2015) and O'Rourke & Walsh 2020 refer to such participants as *Nuachainteoirí* ('new speakers), using the popular Catalonian *linguistic mudes* ap-

[5]RTÉ is Ireland's public-service radio and television network. It manages most of Ireland's radio and television transmitters. Raidió na Gaeltachta is considered one of RTÉ's four major national FM radio channels, and is also streamed live on RTÉ's website.

[6]The Irish language television station, TG4, serves a national audience (meaning mostly L2 speakers), and is mostly not associated with the Gaeltacht, except for being headquartered there.

[7]Belfast was home to a daily Irish language newspaper from the 1980s until 2008, when declining sales for print newspapers nationwide made the venture no longer viable. A popular monthly magazine, *Nós*, still publishes from Belfast's Irish-language center *An Cultúrlann*. An Irish-language insert named *Seachtain* ('Week) still appears weekly in the Dublin-based *Irish Independent*.

[8]There are nascent Irish language communities in Cork City, Galway City, and Derry City, but none has reached the critical mass of speakers that might justify a radio station.

[9]The website https://feilire.com/ lists near-daily language-centered events in and around Galway City, for example.

proach to language and identity. Their central argument is that a speaker's chosen language identity must be respected, regardless of background, and that the long-vaunted "authority of the ideology of the native speaker" is now to be questioned.

Placing these ongoing sociolinguistic debates to one side, the very existence of locally directed radio stations implies the existence of two communities: that of the Gaeltacht, obviously represented in the broadcasters and audience of Raidió na Gaeltachta, and that of the urban areas, also obviously represented in the broadcasters and audiences of Raidió Fáilte and Raidió na Life.[10] Up until now, however, no study has analyzed, compared, or contrasted the Irishes of these two communities. How could one choose informants and control for various factors? What features do the two varieties share? Where do they differ? And what measuring tools could be used to perform such a comparison?

2 Methodology

Gathering data was the first major challenge for this project. In an ideal world, one would find a sufficient number of native speakers from each community, ensuring their approximate comparability in things such as age, gender, socioeconomic background, education, and the like. One would then interview them, under the same circumstances, on a subject that would ensure they would all speak unself-consciously and at length. Alas, in this case such an approach would be impossible without extraordinary resources.

The two communities' broadcasting fora offered an alternative solution. There were three radio stations, all streaming live over the internet, and all using articulate presenters presumably chosen for their comparatively good command of Irish. Studios provided relatively controlled environments, and the variety of programming from each station meant that similar types of programs could be found that allowed for meaningful comparison between speakers. So, Gaeltacht news readers could be compared with urban news readers, and so on with chat show hosts, music presenters, etc.

One caveat in the choice of broadcasters is that those on Raidió na Gaeltachta are mostly professionals, while those on the urban stations are mostly volunteers. It is not clear, however, how or why that might affect the syntax of the

[10]We must, of course acknowledge that Gaeltacht speakers can be heard on the urban stations, as can urban speakers on the Gaeltacht station, but by and large the distinction holds. Raidió na Life has been broadcasting to Dublin since 1993, while Raidió Fáilte has been broadcasting to Belfast since at least 2002.

Irish that they speak, particularly in unscripted situations like chat show monologues. Newsreaders offer good baseline samples of carefully spoken, prepared speech. But news reports are rarely delivered extemporaneously, and it seemed politic to analyze impromptu speech also, also under reasonably controlled circumstances. In this case, I chose presenters of chat shows delivering off-the-cuff monologues at the beginning of their shows. In each case, I chose presenters from the Gaeltacht (on Raidió na Gaeltachta), from Belfast (on Raidió Fáilte), and from Dublin (on Raidió na Life). Additionally, I selected two published texts as upper and lower baselines. These will be discussed further below.

Four speakers from each radio station were originally chosen for the phonetics and morphology study, each group divided between newsreaders and chat show hosts. The total eventually expanded to twenty speakers - ten from the Gaeltacht station, and ten from the two urban stations.

Regarding the speakers, I did as much due diligence as I could to ensure that they were raised in the communities to which they were broadcasting (so, for example, a Dublin speaker had to have been raised within the broadcasting footprint of Raidió na Life; ditto for Belfast speakers; and ditto for Gaeltacht speakers).[11] As noted above, however, it was not possible to control for professional status.

In the case of each broadcaster, I took a 3–5 minute recording and transcribed it. Such recordings usually yielded between 400 and 700 words, or 25–35 sentences, which was more than enough to get a good representation of the speaker's language. My own early work on lexicon suggested to me that c.500 words was enough to approximately calculate a speaker's lexical ability, and this figure has been partially confirmed by the analyses of McCarthy & Jarvis (2007). Other studies suggested that this length of text was enough to also accurately provide sufficient data for syntactic analysis.[12] For syntactic analysis, the total yield of sentences was approximately 660.[13] I then analyzed the recording and the transcription using measuring tools for the categories already described above.

It is worth re-emphasizing that this is not a comparison of native speakers and learners. Using the *linguistic mudes* approach discussed above, I chose my speakers, not on what language they spoke at home, but on what language they were choosing to speak in front of the microphone.

[11]This distinction is not without difficulties. Some broadcasters have been raised in multiple places.

[12]Snowdon et al. (1996) got reliable results from texts that were mostly 200–300 words long.

[13]Twenty speakers with about thirty sentences per transcription, plus the two baseline texts.

Most linguists analyze language using four distinct measures, because they can be systematically observed: Phonetics/phonology, morphology, syntax, and lexicon. I have already published my findings in the areas of and morphology (Ó Broin 2014a), and I do not propose to do any more here than summarize my findings in these areas. A complete phonetic analysis being out of the question, I elected to examine three representative phonetic markers: slender dentals, unvoiced velar fricatives, and unvoiced palatal fricatives.[14] All of these are sounds that are not made in English but are nevertheless necessary for Irish grammar (a slender consonant at the end of a word can indicate plurality, for instance, while a lenited consonant at the beginning of a word can indicate gender).[15] The study revealed a noticeable phonetic variance between Gaeltacht and urban broadcasters, reflected most obviously in expected palatals, of which Gaeltacht broadcasters missed an average of 2%, while their urban counterparts missed an average of 64%.

This would mean little in languages such as English, where phonetics and morphology do not intersect quite so sharply, but the implications for Irish are clear. If speakers are not applying lenition or eclipsis, case forms such as the genitive plural and tense forms such as the past are liable to get lost. As with phonetics, a complete morphological analysis was impossible, and I elected to examine a single representative category - nouns declined in any form. So, for the lemma *cainteoir* ("speaker"), a masculine noun of the third declension, I marked all occurrences of its declined forms: *gcainteoir, cainteora, chainteora, cainteoirí, gcainteoirí,* and *chainteoirí*. I counted as "missed" any instances where the expected change did not occur.

As expected, given the above comments on Irish morphophonology, the phonetic observations in urban broadcast Irish noted above are similarly indexed in morphology. If speakers are not making phonetic changes that mark a genitive or a past tense, these are likely to drop out of the grammatical system. In this case, Gaeltacht broadcasters missed an average of 6% of expected changes, while urban broadcasters missed an average of 45%, rising to 63% in relaxed speech. If city broadcasters are not making 63% of expected morphological changes in relaxed speech, their language is very different from that of Gaeltacht broadcasters. While city broadcasts may sound odd to Gaeltacht speakers, however, Gaeltacht broadcasts must be even more radically different to city speakers, given that Gael-

[14] e.g., the initial consonants of *Dia* ("God") and *teach* ("house"); the initial consonant of *chú* ("dog" [GEN]); and the initial consonant of *cheo* ("fog" [GEN]).

[15] An tendency not to mark plurals, gender or possession suggests that speakers will have great difficulty being understood unless alternatives are already in place.

tacht broadcasters are using sounds not present in the urban phonetic system, as well as a different morphological system.

Next, we turn to the focus of this chapter: Does syntax also vary between these communities? Can qualitative comparison be done in this field? Some work has been done on the general syntax of particular Gaeltacht dialects in Irish (Ó Siadhail 1989, Ó Muirí 1982) but no large-scale comparative studies have been done. While some other categories are worth analyzing here, particularly the presence or absence of Irish-specific features such as the copula or the autonomous form, my principal goal here was to analyze syntactic complexity, which Snowdon et al. 1996 have used as an index of language development.[16] Grammatical complexity, using an approach developed by Sheldon Rosenberg and Leonard Abbeduto in 1987 and refined several times in the following decades, is measurable. Using a scale from 0 to 7, where zero designates a simple monoclausal sentence and seven designates a complex sentence containing embedding of phrases at multiple levels, a text can be analyzed sentence-by-sentence and an average complexity score calculated for the whole text.

Herein lies our first major problem. The Rosenberg and Abbeduto approach is specific to English. How might it be adapted to Irish? With regard to the indexing of Irish and English constructions in the seven-level structure, we can certainly keep very simple monoclausal sentences[17] at level zero and complex ones[18] at level seven, but what about the six levels between them? The revised scale for English awards one point to infinitives or participles, but Irish does not have these in the same manner that English does. Should we similarly score the Irish verbal noun, which is the closest structure in Irish to English non-finite verbal forms? Raising in English is awarded three points, but raising in Irish is very different (McCloskey 1989) – should it receive the same three points in Irish that it receives on the English scale? Verb nominalization receives three or six points in the English scale, depending on whether the nominalization occurs in object or subject position. Is there anything in Irish, whose basic sentence structure is very different (VSO, as opposed to English's SVO), that parallels this? Fortunately for this study, such questions, while important, had little effect on calculations, which ultimately depended very heavily on level-7 sentences.

As we now know, a certain amount of Rosenberg and Abbeduto's classification system initially involved guesswork, hence the several challenges and revisions

[16] Snowdon et al. suggest that "grammatical complexity is associated with working memory, performance on speeded tasks, and writing skill." (Snowdon et al. 1996: 516)

[17] For example, a single proposition containing a single transitive verb.

[18] For example, a sentence containing a tensed or untensed verbal phrase that is nested inside another phrase.

since it came out, and some of this article's approach, for want of extensive data from L1 acquirers of Irish, is similarly open to revision. Some of it comes from the transcriptions that I have made myself of native-speaking children and adolescents (Ó Broin 2014b), but until we have a comprehensive searchable language database for Irish-speaking children, this adaptation will remain, as Rosenberg and Abbeduto's scale did for several decades, a work in progress.

Rosenberg and Abbeduto had been working on measurement tools for ascertaining complexity levels in cognitively disabled adults since 1974, but they only really codified a level-based system in 1987, when they established a seven-level scale (from earliest-to-appear to last-to-appear in normally developing children) for utterance complexity where, for example, a sentence containing an "embedded infinitival complement with subject identical to that of the matrix clause" (e.g., *I am going to meet John*) would receive a single point, while a sentence containing "more than one use of sentence combining in a given sentence" (e.g., *John decided to leave Mary when he heard that she was seeing Mark*) received seven points (Rosenberg & Abbeduto 1987: 26).

This scale was devised using data derived from "literature on the development of complex sentences in normal children" (Rosenberg & Abbeduto 1987: 26). In 1994, Hintat Cheung and Susan Kemper slightly revised this "D-Level" system by adding a zero-level for simple monoclausal sentences, and selected D-Level as one of three reliable measurement tools for the analysis of adult language, particularly in older speakers (Cheung & Kemper 1992: 71). The D-Level scale received a significant boost when it was selected as the grammatical complexity measurement tool in the "Nun Study," (Snowdon et al. 1996: 529) a well-known project involving the analysis of language examples provided by nuns in early life as predictors of mental health or senility in later life. A revision of the scale led by Michael Covington added some new sentence structures and (significantly) swapped levels 5 and 6 (Covington et al. 2006: 10–11), and this revision appears to be the accepted version of the scale since then. It is Covington's revision that I adapt for this analysis (1):

(1) Covington et al. (2006)'s Revision of the Rosenberg and Abbeduto Scale
Level 0 simple sentence
Level 1 non-finite clause as object without overt subject
Level 2 coordinated structure
Level 3 finite clause as object (and equivalents)
Level 4 non-finite clause as object with overt subject (and equivalents)
Level 5 finite or non-finite adjunct clause
Level 6 complex subject
Level 7 more than one structure of Level 1–6

As shown in examples in (2) and (9), defining levels 0 and 7 for Irish is mostly non-controversial. However, scoring sentences between levels 1 and 6 was difficult because Irish has several common grammatical features that are not present in English, as we see in the examples in (3–8). Be it understood, then, that this proposed adaptation of Rosenberg and Abbeduto's scale appears to be mostly reliable but may remain challengeable for specific linguistic features that are found only in Irish. Even so, it is likely that such features would only be promoted or demoted a single level. Be it also understood that this paper is not the appropriate location for an exhaustive description or defense of the adaptation. I intend to provide such a description separately at a later date but do provide some comments here in footnotes.

(2) *Level 0: Simple sentences, including questions.*

a. Bhí ionadaithe ón domhan spóirt i láthair ag
be.PAST representatives from.DEF world sport.GEN in presence at
an Aifreann.
the mass
'Representatives from the world of sport were present at the mass.'

b. Ba leo an lá faoi dheireadh.
COP.PAST to.3PL the day under end
'The day was theirs in the end.' (i.e., They won.)[19]

(3) *Level 1: Simple non-finite verb forms in simple sentences*
Tá sé chun an bus a thógáil.
be.PRS he to the bus to take.VN[20]
'He is going to take the bus.'

[19]There might some argument to be made for giving the copula more complexity, but I have elected to keep simple copular sentences in this category.

[20]There is disagreement among linguists about the labeling of *a* (Lynn 2016: 22–23), and I take Ó Baoill's lead (Ó Baoill 2010: 290) in simply glossing it as "to". In my examples, I mostly stick to the standard Leipzig abbreviations, however a few Irish-specific abbreviations are also needed:

AUT = autonomous form

VA = verbal adjective

VN = verbal noun

(4) *Level 2: Conjoined constructions*
Cosnóidh an t-athnuachan seo ocht gcéad milliún Euro agus
cost.FUT the renovation this eight hundred million Euro and
leanfaidh sé ar feadh trí bliana.
continue.FUT it on duration three years.
'The renovation will cost eight hundred million Euro and it will continue for three years.'

(5) *Level 3: Simple finite subordinate clauses (e.g., relative clauses and noun clauses; subject extraposition)*

a. Dúirt sé nár thaitin an t-amhrán leis.
say.PST he COMP.NEG shine.PST the song with.3SM
'He said that he didn't like the song.'

b. Tá sé tábhachtach go dtiocfaidh gach éinne in am.
be.PRES it important COMP come.FUT every anyone in time.
'It is important that everybody come on time.'

(6) *Level 4: Simple nonfinite verb clauses; autonomous form; stative passive*

a. Thug na húdaráis pleanála cead dóibh
give.PAST the authorities planning.VN.GEN permission to.3PL
athnuachan a dhéanamh ar Bhaile Munna.
renovation to make.VN on Baile Munna.
'The planning authorities gave them permission to renovate Baile Munna.'

b. Bualadh go dona iad.
beat.PAST.AUT badly them.ACC
'They were beaten badly.'

c. Tá an áit tréigthe[21] ag an teaghlach le fada.
be.PRES the place abandoned.VA by the family for long.
'The place was abandoned by the family long ago.'

(7) *Level 5: Subordinate adverbial clauses*
Imreoidh na foirne inniu muna gcuirfidh sé báisteach.
play.FUT the teams today if.NEG put.FUT it rain
'The teams will play today if it doesn't rain.'

[21]It is an open question whether forms like this are verbal adjectives or past participles (Lynn (2016): 62)

(8) *Level 6: Embedded finite or nonfinite clauses in subject position or modifying subject*[22]

a. B' é Jeremiah Roche, a bhí ina bhall de
COP.PAST 3SM Jeremiah Roche REL be.PAST in.3SM.GEN member of
shagairt Chill Téagáin, an fear a maraíodh.
priests Chill Téagáin, the man REL kill.PST.AUT
'It was Jeremiah Roche, who was a member of the Kiltegan Fathers, who was killed.

b. Tá an fear a gortaíodh sa tionósc ag teacht
be.PRS the man REL injure.PAST.AUT in.DEF accident at coming.VN
chuige féin san ospidéal.
to.3SM self in.DEF hospital
'The man who was injured in the accident is recovering in the hospital.'

(9) *Level 7: More than one level of embedding in a single sentence.*
[Deir grúpa caomhnaithe Ghaoth-Barra [go bhfuil imní
say.PRES group preservation.GEN Gaoth-Barra.GEN that be.PRS concern
mhór orthu faoi phleananna an rialtais [ceadúnas a
big on.3PL about plans the government.GEN licence to
eisiúint do chomhlacht mianadóireachta le [dhul i mbun tochailte
issue.VN for company mining.GEN to go.VN about digging.GEN
i gceantair éagsúla i nDún na nGall.]]]
in areas particular in Dún na nGall
'The Gaoth Barra Preservation Group says that they are very concerned about the government's plans to issue a licence to a mining company to begin digging in certain areas of County Donegal.'

Using this scale, I went through every recording in my samples for newsreaders and chat show hosts (plus two benchmark texts, discussed below), giving each of the approximately 660 sentences a D-Level score.

3 Analysis

My earliest approaches to syntactic analysis of these recordings involved taking an entire transcription and calculating the average complexity level for all

[22]For copular constructions like this, I included clauses in either subject position or predicate NP position.

of the sentence-level utterances.[23] This resulted in surprisingly low scores for all recordings, but a reanalysis of the texts demonstrated why: Simple filler sentences and common introductory or salutatory utterances were being included that did not necessarily reflect genuinely spontaneous (or planned) speech.[24] In order to avoid such utterances impacting complexity measurements for speakers, I discarded, after encoding all of the sentences in a transcription, the bottom 50%, an action that predictably caused the complexity scores of all speakers to rise but also led to a more visible separation between speakers whose most complex utterances remained mostly paratactic and speakers whose most complex utterances were hypotactic (see Table 1).[25]

As mentioned above, in addition to the transcribed news reports and chat monologues that I recorded and transcribed, I added two items to establish upper and lower benchmarks. The text for the lower benchmark was a popular idiomatic version of Jack and the Beanstalk written by Mairéad Ní Ghráda for children (Ní Ghráda 2002), which I was hoping would reflect the lowest realistic score that a coherent text might get, and the second a passage from Máirtín Ó Cadhain's short story *Cé Acu* (Ó Cadhain 2009), which I think represents a very complex text.[26] Applying the same "top 50%" approach, the children's story scored 2.9[27], while the Ó Cadhain text scored 6.9.[28]

3.1 All recordings

The D-Level scores for the top 50% of sentences in all twenty recordings and the two benchmark texts are shown in Table 2.

[23] Coding such texts is very time-consuming, requiring exact transcription of relatively long texts, transfer to a spreadsheet, complexity-coding for every utterance, and calculation of averages for the entire text. For this reason, the sample-size in this study is small, but big enough to be provisionally representative, as the statistical analysis in this paper demonstrates.

[24] E.g. *Tá sé cúig nóiméad tar éis a haon.* 'It is five minutes past one. *Bhur gcéad fáilte isteach inniu* 'You're [all] very welcome to [the show] today'; *Sin a bhfuil ón seomra nuachta* 'That's it from the newsroom'; etc.

[25] I.e. utterances in which the clauses are presented in parallel, using coordinating conjunctions, versus utterances in which the clauses are presented as superordinate and subordinate by the use of subordinating conjunctions.

[26] One might be able to find more complex texts using, for example, legal documents or acts of the Oireachtas, but the language of such texts is highly artificial and unsuited for comparison with language intended for real-time communication.

[27] Including the bottom 50% of utterances, this text had an average score of 0.9. One could make an argument for using this lower score (since it reflects a coherent text), but it seems only fair to treat this text in the same way that the recordings were treated.

[28] Including all utterances, this fell to 4.35.

Table 1: Sample transcript with complexity score for top 50% of sentences

#	Sentence	Compl.
1	Na cinnlínte nuachta (ar) raidió na life.	0
2	Is mise XXX XXX	0
3	D'fhreastail na céadta daoine ar shochraid an churaidh oilimpeach Darren Sutherland san Uaimh i gContae na Mí um thráthnóna.	0
4	Bhí ionadaithe ón domhan spóirt i láthair ag an Aifreann.	0
7	Thángadar ar chorp Darren ina árasán i Londain an Luain seo caite.	0
15	Beidh an fear ag teacht os comhair cúirte i gClover Hill an ochtú lá is fiche den mhí seo.	0
16	"Bhuaigh an t-aisteoir Brendan Gleeson Emmy aréir i Los Angeles ar son a ról mar Winston Churchill sa scannán ""Into the Storm""."	0
18	Agus anois an aimsir.	0
19	Lá réasúnta séimh inniu le seal grian anois is arís.	0
20	Beidh beagáinín báisteach ag teacht ón iarthar i dtreo an oirthir níos déanaí inniu.	0
21	An teocht is airde idir cúig déag agus ocht déag gcéim celsius.	0
22	Agus sin deireadh na nuachta go fóill.	0
10	Cosnóidh an t-athnuachan seo ocht gcéad milliún Euro agus leanfaidh sé ar feadh trí bliana, ag giniúint dhá mhíle post.	2
12	Faoin scéim seo beidh trí chéad is a fiche árasán, leabharlann poiblí, pictiúrlann, oifigí, agus ionad siopadóireachta.	2
17	Bhí sé i gcoinne aisteoirí den scoth ar nós Sir Ian McKellen agus Kiefer Sutherland.	2
5	In áirithe bhí Kenny Egan, céile imeartha Darren, a chaith a chulaith Éireannach in ómós do Darren.	3
11	Tá an ceadúnas tógála seo ar cheann de na ceadúnais is mó a tugadh cead dóibh i mBaile Átha Cliath.	4
6	Cuimhníodh ar Darren mar laoch spóirtiúil ach freisin mar fhear mór muinteartha.	4
8	Creidtear [gur chuir sé lámh ina bhás féin].	4
9	Tá lánchead pleanála faighte ag Treasury Holdings inniu <chun athnuachan a dhéanamh ar Bhaile Munna>.	4
14	Fuair(eadh) corp Lisa Doyle, fiche ceathar, i gContae Ceatharlach maidin inné.	4
13	[Tá fear tríocha haon bliain d'aois tar éis teacht os comhair cúirte inniu maidir le bás bean i gcontae ceatharlach.]	4
		33
		3.0

Table 2: All D-level scores per recording

#		Gaeltacht or City	Genre	Top 50% Complexity
9	Gaeltacht Chat	Gaeltacht	Chat	2.3
22	SeanGasPonaire	Yardstick		2.9
5	Urban RnL Nuacht	City	News	3
4	Urban RnL Chat	City	Chat	3.23
34	Urban RF Chat	City	Chat	3.69
6	Urban RF Chat	City	Chat	3.86
28	RnL Nuacht	City	News	4.07
25	Urban RF Nuacht	City	News	4.4
1	Urban RF Nuacht	City	News	4.45
35	Urban Nuacht RnL	City	News	5
38	Gaeltacht Chat	Gaeltacht	Chat	5.1
10	Urban RnL Chat	City	Chat	5.22
27	Urban RnL Chat	City	Chat	5.27
2	Gaeltacht Chat	Gaeltacht	Chat	5.66
36	Gaeltacht Chat	Gaeltacht	Chat	5.77
41	Gaeltacht Chat	Gaeltacht	Chat	5.78
26	Gaeltacht Nuacht	Gaeltacht	News	6
33	Gaeltacht Nuacht	Gaeltacht	News	6.27
3	Gaeltacht Nuacht	Gaeltacht	News	6.28
31	Gaeltacht Chat	Gaeltacht	Chat	6.83
21	Máirtín Ó Cadhain	Yardstick		6.9
43	Gaeltacht Nuacht	Gaeltacht	News	7

There is an obvious distinction to be seen immediately. Excluding the benchmark texts (labeled Yardstick), the eight highest-scoring items, those with complexity scores between 5.66 and 7.0, are Gaeltacht recordings. Conversely, ten of the twelve lowest-scoring items, those with scores between 3.0 and 5.27, are urban recordings. The urban average is 4.22, while the Gaeltacht average is 5.70. This suggests very strongly that there is a measurable dissimilarity between the syntactic complexity of Gaeltacht broadcast Irish and the syntactic complexity of urban broadcast Irish.

3.2 Statistical analysis

While Table 2 looks persuasive, it seemed politic to perform a statistical analysis on the data to confirm the significance of the findings. A power analysis,

performed using SPSS, returned a score of 0.86 (that is, an 86% chance that the design of the study was reliable), a very healthy score for a small-scale study like this. For significance, I applied an Independent Means t-test to the complexity scores of the broadcasters, treating urban broadcasters as sample 1 and treating Gaeltacht broadcasters as sample 2. The null hypothesis was that there was no significant difference between sample 1 and sample 2, and the alternative hypothesis was that there was a statistically significant variance in syntactic complexity between the two groups. My calculations, performed both by hand and using SPSS, returned a p-value of 0.008 (equal variances not assumed), well below the standard significance threshold of 0.05, and arguing very strongly against the null hypothesis.

3.3 Limitations

There are three limitations to this study. The first is sample size (ten Gaeltacht broadcasters and ten urban broadcasters), but both the power analysis and the t-test figures strongly suggest that the study is reliable and that there is a statistically significant variance between the two samples.

The second limitation is in the measurement of complexity scores. That is to say, observer bias in the scoring of the sentences cannot be discounted. A well-funded project involving a widely spoken language would be able to train and pay two individuals to independently assign complexity scores to the c.660 discrete sentences collected from all the recordings, thus ensuring a minimum of unconscious bias during the assignment of complexity scores. Alas, for a minority language like Irish, this was simply out of the question. Those individuals in the world who are comfortably conversant with syntactic theory, the Irish language, and Covington's adaptation of Rosenberg and Abbeduto's complexity score could be optimistically counted on a single hand.

The third limitation is the professional status of the broadcasters. It may well be that professional status affects the complexity score of a broadcaster, and for that reason, this study's findings remain localized to the broadcasters on the three stations.

The above analysis shows a clear distinction between the two samples, but it also treats all the recordings as equal. However, it is not a surprising admission to acknowledge that there is a difference between the prepared texts of news recordings and the ad-hoc nature of chat show monologues. This study, therefore, will analyze the two genres separately.

3.4 Comparison of news broadcasts

Since newsreaders read from prepared texts that describe often complex situations, one can assume that the language of their reports reflects careful thought and comparatively complex language. Material will be similar, whether the newsreader is working from an urban or Gaeltacht newsroom; audiences will be similar (educated adults); the speakers' situations will be similar: they have prepared and printed out a news report beforehand and are reading it aloud live in a studio.

Table 3: Low-high complexity scores for newsreaders

#		Gaeltacht or City	Genre	Top 50% Complexity
22	SeanGasPonaire	Yardstick		2.9
5	Urban RnL Nuacht	City	News	3
28	RnL Nuacht	City	News	4.07
25	Urban RF Nuacht	City	News	4.4
1	Urban RF Nuacht	City	News	4.45
35	Urban Nuacht RnL	City	News	5
26	Gaeltacht Nuacht	Gaeltacht	News	6
33	Gaeltacht Nuacht	Gaeltacht	News	6.27
33	Gaeltacht Nuacht	Gaeltacht	News	6.28
21	Máirtín Ó Cadhain	Yardstick		6.9
43	Gaeltacht Nuacht	Gaeltacht	News	7

The average complexity score for urban newsreaders is 4.18, while that of Gaeltacht newsreaders is 6.39, and while the number of recordings for this study is comparatively small, there is a very clear pattern visible. As shown in Table 3, no Gaeltacht newsreader scored less than 6.0, while no urban newsreader scored higher than 5.0.

The urban average may suggest that urban newsreaders are making sentences that contain basic adjectival and nominal clauses (level 4), but the real absence is of level-seven sentences in urban news reports, that is, sentences containing complex, nested clauses such as in (10).

(10) [Agus deir Cónaidhm Árachais na-hÉireann anois go mbeidh
and say.PRS federation Insurance.GEN of-Ireland now that be.FUT
méadú ar tháillí árachais de-bharr an mhéid airgid (a
increase on costs insurance.GEN because the sum money.GEN REL

íocadh le úinéirí tithe le ceithre mhí)].
pay.PAST.AUT to owners houses.GEN for four months
'And the Insurance Federation of Ireland now says that there will be an increase in insurance charges because of the amount of money that was paid to house owners in the last four months.'

The urban strategy for forming complex sentences seems to involve stringing together a series of utterances and connecting them with conjunctions, as shown in (11). The strategy works, but makes for sentences that require patience from the listener to connect the various ideas together:

(11) Tá práinn tagtha ar an dóigh a bhfuil cúrsaí anois
be.PRS urgency come.VA on the way REL be.PRS events now
de-thairbhe neamhshocracht ins na margaí mar-gheall-ar an ngníomh
because instability in the markets because the action
seo, agus fosta mar-gheall-ar seo tá laghdú tagtha inniu ar
this and also because this be.PRS fall come.VA today on
chostas na n-iasachtaí fá-choinne gheilleagair na-hEorpa, agus
cost the loans.GEN for economy Europe.GEN, and
mar-gheall-ar an ngéarchéim chéanna fosta, tá cruinniú socraithe
because the crisis same also, be.PRS meeting arrange.VA
ag an aire airgeadais Sammy Wilson anseo chun éifeacht na
by the minister finance.GEN Sammy Wilson here for effect the
géarchéime ar an ghéilleagar seo a mheas.
crisis.GEN on the economy this to measure.VN
'Urgency has come to the way things are because of instability in the markets because of this action, and also because of this the cost of loans has fallen for the European economy, and because of the same crisis also, the finance minister Sammy Wilson has arranged a meeting to ascertain the effect of the crisis on the economy.' [29]

Note the outlier in this group, who is an urban newsreader returning a D-Level score of 3.0. The speaker nevertheless does not seem to have any particularly idiosyncratic linguistic features that stand out. Their logical strategy seems to simply involve using pronouns as the principal aid in connecting one sentence to the next, as we see in (12):

[29] I have attempted to retain the manner in which the original Irish's clauses *squint*. That is, the listener is not certain what relationship they are in with the preceding and foregoing clauses.

(12) Bhuaigh an t-aisteoir Brendan Gleeson Emmy aréir i Los Angeles
win.PAST the actor Brendan Gleeson Emmy last-night in Los Angeles
ar-son a ról mar Winston Churchill sa scannán "Into the Storm".
for his role as Winston Churchill in-the film Into the Storm.
Bhí sé i-gcoinne aisteoirí den-scoth ar-nós Sir Ian McKellen agus
be.PAST he against actors excellent like Sir Ian McKellen and
Kiefer Sutherland.
Kiefer Sutherland.
'The actor Brendan Gleeson won an Emmy last night in Los Angeles for his role as Winston Churchill in the movie "Into the Storm." He was [up] against excellent actors like Sir Ian McKellen and Kiefer Sutherland.'

This material might possibly be compressed into a single, more-complex sentence by a speaker with higher D-Level averages.[30]

Equally interesting are the very stable figures returned by Gaeltacht newsreaders, whose reports all achieved complexity scores between 6.0 and 7.0, suggesting both a very high frequency of level-seven sentences, and a very stable level of complexity. Urban syntactic complexity swung between 3.0 and 5.0. Much as the figures for phonetics and morphology suggested, the syntactic complexity of Gaeltacht news texts is predictable and stable, while that of urban newsreaders varies widely.

3.5 Comparison of chat show monologues

Chat shows offered a better view of extempore language. The broadcasters were still speaking under controlled studio conditions, but apart from the possibility of having notes to remind them of what was coming up in their shows, it was virtually certain that their monologues were unscripted and that their language therefore reflected the genuine day-to-day patterns of the speakers.

A similar dichotomy exhibited itself with these speakers: the D-Level average for urban chat show hosts was 4.25, while that for Gaeltacht chat show hosts was 5.24. As shown in Table 4, no urban broadcaster had a higher score than 5.27.

Level five of the Rosenberg and Abbeduto scale marks a level at which speakers are forming relatively complex adverbial clauses. That is to say, they are creating relationships between two separate events at main-clause and subordinate-clause level. Level four mostly concerns itself with structures that involve non-

[30] For example, by subordinating the second sentence in the following manner: "Fighting off stiff competition from a star-studded list of candidates that included Sir Ian McKellen and Kiefer Sutherland, ...".

Table 4: Complexity scores for chat show hosts

#		Gaeltacht or City	Genre	Top 50% Complexity
9	Gaeltacht Chat	Gaeltacht	Chat	2.3
22	SeanGasPonaire	Yardstick		2.9
4	Urban RnL Chat	City	Chat	3.23
34	Urban RF Chat	City	Chat	3.69
6	Urban RF Chat	City	Chat	3.86
38	Gaeltacht Chat	Gaeltacht	Chat	5.1
10	Urban RnL Chat	City	Chat	5.22
27	Urban RnL Chat	City	Chat	5.27
2	Gaeltacht Chat	Gaeltacht	Chat	5.66
36	Gaeltacht Chat	Gaeltacht	Chat	5.77
41	Gaeltacht Chat	Gaeltacht	Chat	5.78
31	Gaeltacht Chat	Gaeltacht	Chat	6.83
21	Máirtín Ó Cadhain	Yardstick		6.9

finite verbal structures, but not necessarily the establishment of relationships between events, as we find in the subordinate clauses of level 5.

The average for urban broadcasters falls well below level five, suggesting that adverbial clauses are uncommon in their speech. Three of the five urban broadcasters return scores below level four, suggesting that they rarely form sentences that include subordinate non-finite verb structures. Regarding the creation of relationships between utterances (which a complex speaker would achieve through the use of subordinating conjunctions and/or nested clauses), urban broadcasters appear to be expressing such relationships by using coordinating conjunctions and expecting that the listeners will intuit the relationships themselves, as in (13).

(13) So, [tháinig sé ar-ais óna laethanta-saoire] **agus** [bhí a
so come.PAST he back from-his holidays and be.PAST his
dhoras chairr briste síos] **agus** [bhí, bhí a chlann tar-éis,
door car.GEN broken.VA down and be.PAST be.PAST his family after
sochraid **agus**, you know, mar sin,[a bheadh orthu]], **agus**, [bhí
funeral and you know therefore REL be.COND on-them and be.PAST
gach, gach éinne, like, "Tá tú marbh" [nuair a tháinig sé
every every person like be.PRS you dead when REL come.PAST he

ar-ais]].
back
'So, he came back from his vacation and his car door was broken down and his family were after having have a funeral, you know, and, therefore, and everyone was like "you're dead" when he came back.'

There is nevertheless one obvious outlier among the chat show hosts, just as there was among the newsreaders: the lowest score of all was from a Gaeltacht broadcaster, who returned a score of 2.30. This presenter, a popular and well-known Gaeltacht personality, has a very recognizable telegrammatic style involving clipped single-phrase utterances that he frequently repeats. Despite his frequent use of conjunctions, it was usually clear when a new utterance was beginning, however, and even if more than two utterances could be conjoined, the resulting utterance would still only reach a D-level complexity of 2.

One would expect the D-level of news reports to be measurably higher than that of extemporaneous chat show monologues, and that is what happened in the Gaeltacht, where the newsreaders' average was 6.39 and the chat shows' average was 5.24. Extemporaneous speech does not lend itself so readily to complex, planned sentences containing deeply nested clauses and phrases. It is perhaps surprising, therefore, to find that the D-Level average for urban chat show hosts (4.25) is *higher* than that of urban newsreaders (4.18). If we do allow the possibility that urban broadcasters, whether in prepared or extemporaneous speech situations, typically hover around the 4.2 mark, and if we recall that complex level-7 sentences appear to be making the most difference in these calculations, this suggests that urban broadcasters do not form such sentences, even when given ample time to prepare.

4 Conclusions

In prepared texts, there is a clear difference between the complexity scores for Gaeltacht news reports versus those of urban news reports. The Gaeltacht average (6.39) was more than two full points ahead of the urban average (4.18) on the eight-category D-Level scale adapted from Rosenberg and Abbeduto. This is a significant finding, and suggests that urban broadcast Irish, even in non-extemporaneous conditions where the speakers had ample time to prepare their texts, is being prepared on a complexity level significantly below that of the Gaeltacht and below that at which sentences with adverbial clauses are usually being made. Given that in previous research I had already found major differences in

pronunciation and morphology, this suggests that urban broadcast Irish is syntactically very different from Gaeltacht broadcast Irish. In extemporaneous speaking situations, the average complexity level of Gaeltacht chat show hosts did fall one point (to 5.24), and this is what we would be expecting, given that complex, level-7 sentences often take preparation time that is not available to speakers who are planning sentences on the fly. This figure nevertheless remained nearly a point ahead of urban extemporaneous language. The most significant finding here, however, was that urban figures hardly moved. In fact, the urban average actually rose fractionally, from 4.18 to 4.25. This strongly suggests that urban broadcasters, whether given time to prepare their sentences or not, are mostly not forming complex sentences with nested phrases.

The figures for newsreaders provide a good illustration of how a D-Level of 5 seems to be the point at which the urban and Gaeltacht broadcasters separate. That is to say, level 5 sentences (e.g. subordinate adverbial clauses) are occasionally being formed by urban broadcasters, but level 7 sentences (complex multiclausal sentences with nested phrases) are mostly not. Gaeltacht broadcasters were much quicker to form complex sentences of this sort. Indeed, 72% of Gaeltacht newsreader's sentences were scored at level 7, while only 15% of urban sentences received similar scores.[31]

Staying with D-7 sentences, chat show hosts reflected the pattern that we saw above in our analysis of total D-Level figures. Gaeltacht extemporaneous speech D-7 figures fell from 72% to 33%, but urban figures hardly moved (rising from 12% to 15%). This suggests that urban broadcasters, whether or not they have time to prepare sentences, tend to produce only about 12%-15% of their most complex sentences at D-7 level.

One must note that languages encode complexity in different ways. An absence of embedded verbal clauses does not necessarily imply that a language or dialect is somehow less complicated. Nor does the frequent presence of embedded verbal clauses suggest that a language or dialect is necessarily more sophisticated. One would, however, expect equal levels of parataxis and hypotaxis between dialects of the same language, and it is therefore noteworthy that the syntactic complexity of urban and Gaeltacht broadcasters (as defined using the adapted Rosenberg and Abbeduto scale) is significantly different. That is, urban and Gaeltacht Broadcasters of Irish appear to have measurably different expectations of their listeners with regard to the manner in which they connect clauses.

[31]These percentages are for each recording's top half of sentences, of course, and perhaps could be halved to represent the total of level-7 sentences for each speaker.

5 Implications and further work

It remains an open question, and now one in urgent need of addressing, whether urban broadcast Irish is syntactically simpler because its speakers are L2 speakers. If measurable differences can be found between urban L1 and L2 speakers, this would bolster the suggestion that urban Irish is unlikely to develop in syntactic complexity unless it is adopted by parents as a home language and accrues more native speakers.

The sample for this study was relatively small: twenty speakers in total. No attempt was made to control for gender, age, education, professional status, or (crucially) status as an L1 or L2 speaker. The only control was for approximate geographical origin. So, due diligence was done in ensuring that a Dublin speaker was, indeed, a Dublin speaker, a Belfast speaker was a Belfast speaker, and a Gaeltacht speaker was, indeed, a Gaeltacht speaker. As noted above, two other limitations to this study must be acknowledged: that the professional/amateur status of the Gaeltacht/urban broadcasters may have influenced their syntax scores; and that unconscious bias on the scorer's part may have influenced the D-level scoring of sentences. For this reason, it is best to view this paper's findings on the differences between urban and Gaeltacht Irish as preliminary and subject to confirmation or rejection in future studies.

As I noted in my previous work on phonetics and morphology, urban broadcasters tend to be discarding features of Irish that are not found in English. Velar and palatal fricatives are being dropped in favor of the nearest English sound, for example, while nouns are frequently no longer morphophonetically marked for case, with eclipsis and lenition becoming optional. In this paper we make a compelling case that urban broadcast Irish is also significantly different in syntax, substituting subordinators with conjunctions that require the listener to intuit the relationships between clauses and rarely forming sentences that involve the nesting of embedded phrases and clauses.

Acknowledgements

I would like to extend my sincere thanks to several scholars: Andrew Carnie and the two anonymous readers of an early draft of this paper for suggesting a more formal statistical analysis; Michael Gordon of the Psychology Department, WPUNJ for guiding me in the direction of power analysis; and Nicole Vittoz of Douglas College, British Columbia, for providing procedural guidance in this particular analysis. Any errors, of course, are my own.

References

Armstrong, Timothy. 2012. Establishing new norms of language use: The circulation of linguistic ideology in three new Irish-language communities. *Language Policy* 11. 145–168. DOI: 10.1007/s10993-011-9219-2.

Cheung, Hintat & Susan Kemper. 1992. Competing complexity metrics and adults' production of complex sentences. *Applied Psycholinguistics* 13. 53–76. DOI: 10.1017/S0142716400005427.

Covington, Michael, Congzhou He, Cati Brown, Lorina Naçi & John Brown. 2006. *How complex is that sentence? A proposed revision of the Rosenberg and Abbedulto D-level scale*. Tech. rep. University of Georgia Artificial Intelligence Center. https://ai1.ai.uga.edu/caspr/2006-01-Covington.pdf (5 June, 2025).

Edwards, John. 2016. Minority languages and group identity. In Siân Preece (ed.), *The Routledge handbook of language and identity*, 492–503. London: Routledge.

Gaeloideachas. 2023. *Statistics*. https://gaeloideachas.ie/i-am-a-researcher/statistics/ (28 October, 2023).

Lynn, Teresa. 2016. *Irish dependency treebanking and parsing*. Dublin City University. (Doctoral dissertation).

McCarthy, Philip M. & Scott Jarvis. 2007. Vocd: A theoretical and empirical evaluation. *Language Testing* 24. 459–488. DOI: 10.1177/0265532207080767.

McCloskey, James. 1989. Is there raising in Modern Irish? *Ériu* 31. 59–99.

Ní Ghráda, Mairéad. 2002. *Seán agus an gas pónaire*. Baile Átha Cliath: An Gúm.

Ó Baoill, Dónall P. 2010. Mood in Irish. In Björn Rothstein & Rolf Thieroff (eds.), *Mood in the languages of Europe*, 273–291. John Benjamins: John Benjamins. DOI: 10.1075/slcs.120.15oba.

Ó Béarra, Feargal. 2007. Late Modern Irish and the dynamics of language change and language death. In Hildegard Tristram (ed.), *The Celtic languages in contact: Papers from the workshop within the framework of the XIII international congress of Celtic studies*. Potsdam: Universitätsverlag Potsdam. https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/deliver/index/docId/1750/file/260_269.pdf (5 June, 2025).

Ó Broin, Brian. 2014a. New urban Irish: Pidgin, creole, or bona fide dialect? the phonetics and morphology of city and Gaeltacht speakers systematically compared. *Journal of Celtic Linguistics* 10. 69–91.

Ó Broin, Brian. 2014b. Ord insealbhaithe deilbhíochta agus comhréire Gaeilge i naíonáin. In *Teangeolaíocht na Gaeilge xiii*, 65–85. Uppsala: Uppsala Universitet.

Ó Cadhain, Máirtín. 2009. Cé acu. In *An tSraith ar lár*, 143–186. Indreabhán: Cló Iar-Chonnacht.

Ó Cinnseala, Colm. 2020. *Plean teanga: limistéar pleanála teanga eachréidh na gaillimhe*. https://udaras.ie/assets/uploads/2020/05/20-An-tEachreidh.pdf (5 June, 2025).

Ó Gadhra, Nollaig. 1988. Irish government policy and political development of the gaeltacht. *Language, Culture and Curriculum* 1(3). 251–261. DOI: 10.1080/07908318809525044.

Ó Muirí, Damien. 1982. *Comhréir Ghaeilge Ghaoth Dobhair: Pátrúin na n-abairtí*. Baile Átha Cliath: Coiscéim.

Ó Siadhail, Mícheál. 1989. *Modern Irish: Grammatical structure and dialectal variation*. Cambridge: Cambridge University Press.

O'Rourke, Bernadette & John Walsh. 2020. *New speakers of Irish in the global context: New revival?* New York: Routledge.

Rosenberg, Sheldon & Leonard Abbeduto. 1987. Indicators of linguistic competence in the peer group conversational behavior of mildly retarded adults. *Applied Psycholinguistics* 8(1). 19–32. DOI: 10.1017/S0142716400000047.

Snowdon, David, Lydia Greiner, David Wekstein, Deborah Danner, William Markesber, Susan Kemper & James Mortimer. 1996. Linguistic ability in early life and cognitive function and alzheimer's disease in late life. findings from the nun study. *Journal of the American Medical Association (JAMA)* 275(24). 528–32. DOI: 10.1001/jama.1996.03530480021021.

Walsh, John & Bernadette O'Rourke. 2015. *Mudes* teangeolaíocha agus nuachainteoirí na Gaeilge. *Comhar Taighde* 1. 2–14. DOI: 10.18669/ct.2015.09.

Watson, Iarfhlaith. 2002. Irish-language broadcasting: History, ideology and identity. *Media, Culture, and Society* 24(1). 739–757. DOI: 10.1177/016344370202400601.

Chapter 10

Bunadas: A network database of cognate words, with emphasis on Celtic

Caoimhín P. Ó Donnaíle
Sabhal Mòr Ostaig

Bunadas is a network database of cognate words, freely available as a WWW resource. It contains so far about 100,000 words in about 40 Indo-European languages ancient and modern, going back to PIE itself, and including 70,000 in the Celtic languages. Cognate words are useful for language learning, especially vocabulary. They provide insight into the basic underlying meanings of words, and into the history of languages, peoples and technologies. Etymological dictionaries typically provide long, but rather random and very incomplete lists of cognates, with lots of inefficient duplication between dictionaries, and small languages often lack any good etymological dictionary at all. It is more efficient to assemble together the information for many languages at once. However, the tabular, rectangular structure typically used for short illustrative lists of cognates simply does not scale up. Bunadas uses instead a network structure, which enables "genealogical trees" of cognates to be generated on the fly. This chapter describes some vital aspects of its design and gives some thoughts on its future.

1 Introduction

Cognate words provide insight into the basic underlying meanings of words, and into the history of languages and peoples, thoughts, and technologies. Etymological dictionaries typically provide long, but rather random and very incomplete lists of cognates, with lots of inefficient duplication between dictionaries. This chapter describes a new online tool for researchers developed by the author, that allows for comparative analysis of lexemes across Indo-European languages. I explore the structure and functioning of the database and search tool. Using Celtic

Caoimhín P. Ó Donnaíle. 2025. Bunadas: A network database of cognate words, with emphasis on Celtic. In Andrew Carnie, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond & Luis Irizarry (eds.), *Foundational approaches to Celtic linguistics*, 369–382. Berlin: Language Science Press. DOI: 10.5281/zenodo.15654875

as a test case, I demonstrate how an output based on trees that can be generated "on the fly" provides more insightful etymological results than traditional tabular comparison tools.[1]

Figure 1 shows an example of the kind of cognate tree that Bunadas can generate on the fly: the tree starting from the Proto-Indo-European word *ḱwṓ* 'dog' and showing descendants (reflexes) such as Proto-Celtic celt:*kū*, Old Irish sga:*cú*, Scottish Gaelic gd:*cù*, Proto-Germanic germ:*hundaz*, Old English ang:*hund*$^{\text{dog}}$, and English en:*hound*. Each word in Bunadas has two, or occasionally three, components. First comes a standard[2] language tag such as sga (Old Irish) or gd (Scottish Gaelic). Important for usability, this is color-coded to reflect the language family: green for Celtic, black for Germanic, red for Romance, blue for Slavic, bronze for Greek. Next comes the actual spelling of the word, and again this is color-coded to reflect finer distinctions. The general progression is from white or gray colors for the proto languages, progressing to bolder colors for the modern languages.

Finally, there can be a third component, a disambiguator to distinguish homographs. We see this in the Old English word ang:*hund*^{dog} for example, which has been given the disambiguator 'dog', because there is another word ang:*hund*^{100} meaning 'hundred'. For languages which have a single very authoritative dictionary, such as Welsh (GPC[3]) (Thomas et al. 2022) or Old Irish (DIL[4]) (Toner et al. 2019), Bunadas uses the disambiguator from this dictionary, but in other cases it can use any arbitrary text string which will serve the purpose. This disambiguation of homographs is vital to make Bunadas function. Internally, Bunadas works using numeric IDs for words and is therefore not confused by homographs, but these ID tags are not seen by users so these disambiguatordisambiguators are critical for users.

2 Words and clusters

The trees that Bunadas displays are perhaps its most impressive feature, but Bunadas does not in fact store anything in tree form. Its basic building blocks are words and clusters of words. Using the word-cluster-word connections, it connects the words into a network, a network which is not a tree, and which can

[1]This chapter is best read while simultaneously trying out the online resource itself at https://www.smo.uhi.ac.uk/teanga/bunadas.

[2]ISO 639-1 where available, or ISO 639-3, with occasional ad hoc additions for ancient and proto languages.

[3]https://www.geiriadur.ac.uk/

[4]https://dil.ie/

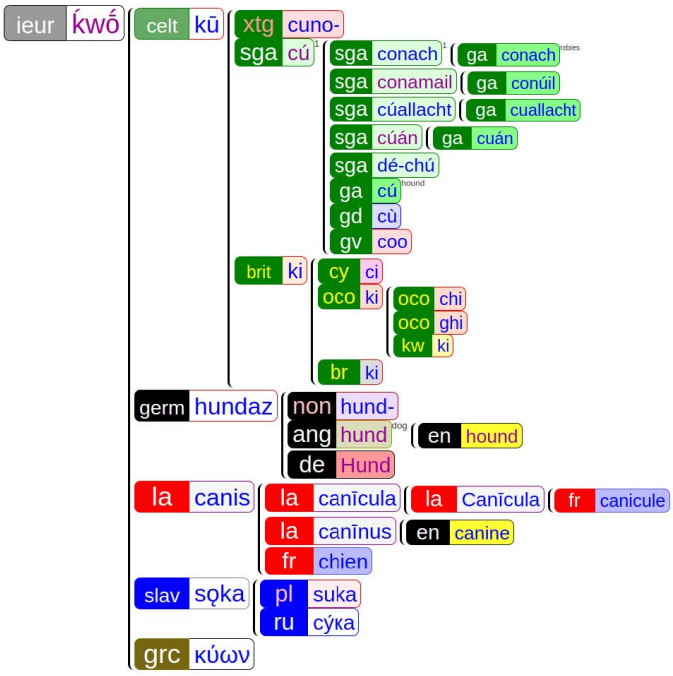


Figure 1: A tree starting from the Proto-Indo-European word *ḱwṓ*

and does contain lots of cycles.[5] When the user requests, though, to see a tree starting from say ieur:*ḱwṓ*, Bunadas walks the network and produces a minimal spanning-tree of words cognate to *ḱwṓ*.

A cluster in Bunadas is just a particular group of words which are all related to one another etymologically. This includes not only groups of words from different languages, but also groups of morphological derivatives within a language. Clusters can overlap, and they can be used in whatever way the compiler(s) and editor(s) of Bunadas find convenient.

If one clicks on the word ang:*hund*^{dog}, this produces the webpage shown in Figure 2. At the very top of the page, there is a fuzzy M on a yellow background. This connects to a dictionary lookup of the word via Multidict.[6] Multidict (Ó Donnaíle 2014) contains a database of information on lots of online dictionaries,

[5] A "cycle" just means that you can follow the network round in a circle, such as germ:hundaz → de:*Hund* → en:*hound* → germ:*hundaz* (Proto-Germanic to German, to English, back to Proto-Germanic).

[6] https://multidict.net

including etymological dictionaries, in lots of languages, together with the parameters which enable it to link through to them transparently. To the right of the yellow M, there is a [W] favicon which connects to a Wiktionary[7] lookup of the word. Thus, we see that Bunadas affords excellent connections to other resources. Following that at the top of the page are the properties stored for the word ang:*hund*^{dog} in the Bunadas database. At the very bottom of the page is a grey section "Lexically similar words", showing hypergraphs, homographs and hypographs[8] (if any), but this is just additional information which might be of use to users and editors and is nothing to do with Bunadas itself.

The section in the middle, showing the clusters, is what is important for the structure of Bunadas. We see that our word ang:*hund*^{dog} belongs to three clusters. The first connects it with its modern English reflex en:*hound*. The second connects it with another Old English word ang:*grīġhund*. The third puts it in a cluster of words descended from Proto-Germanic germ:*hundaz*.

The numbers shown after the words in the clusters, are a metric indicating the distance of the word from the center of the cluster. Normally, we try to pick some word to be at the center of the cluster, at distance 0. This would normally be the word in the proto language if all the other words are descendants of it; or the basic stem word if the others are morphologically related. It is not necessary to have a center word. A cluster of cognate words with no known antecedent could show them all as distance 1 from the cluster, and they would then each be distance 1 + 1 = 2 from each other. The metric is not restricted to values 0 or 1. It can be any number at all, such as 2, or 4.3, and this can be useful when we know that two words are related, but probably rather distantly. Bunadas is not normally very sensitive to the distance metric because the important thing is the connectivity, so for the most part the metric is 1. However, tweaking this metric a bit, from 1 to 1.1 or to 0.9, can be useful when we want to refine the tree to show word derivation via a particular path, such as sga:*cú* → sga:*cúallacht* → ga:*cuallacht* (synchronic in Old Irish, followed by diachronic to Modern Irish), in preference to sga:*cú* → ga:*cú* → ga:*cuallacht* (diachronic to Irish, followed by synchronic in Modern Irish).

Note that the way the trees are constructed, as minimal spanning trees calculated by walking through the network and minimizing the total metric distance, means that if a cluster was duplicated, there would be no difference whatsoever in the trees. This was a design decision early on in the design of Bunadas. An

[7] https://en.wiktionary.org/

[8] Homographs are words which are identical in the way they are written, such as "invalid" (disabled person) and "invalid" (not valid), while "invalidity" would be a hypergraph and "valid" a hypograph.

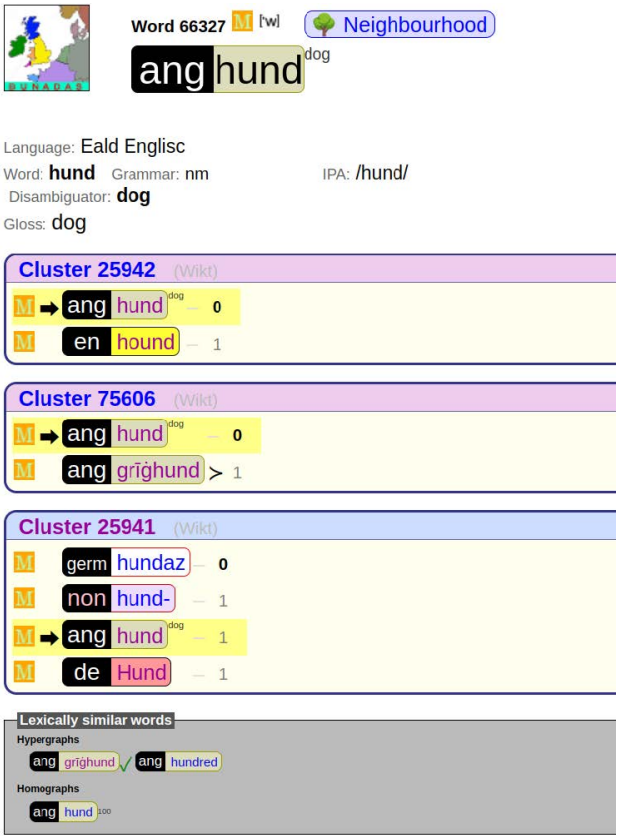


Figure 2: The word ang:hund^{dog} showing the clusters it belongs to.

alternative approach would be to strengthen connection among words if they are related in two different clusters, perhaps derived from two different information sources. Rejecting this alternative approach and choosing to use minimal spanning trees has proven in practice to be the right decision.

In the second cluster shown in Figure 2, ang:*hund*^{dog} is linked to ang:*grīġhund*. We might ask why ang:*grīġhund* does not appear in the tree shown in Figure 1. The solution lies in the ≻ symbol shown after *grīġhund* in the cluster. This symbol means that *grīġhund* is made up of two (or more) main components, and *hund* is only one of them. In order to see ang:*grīġhund* in the tree, it is necessary to switch on the option "Show super-elements". The result is shown in Figure 3.

Figure 3 also illustrates some other features of Bunadas. Hovering over a word causes the part of speech and gloss for the word to pop up, in this case "[nm] greyhound".

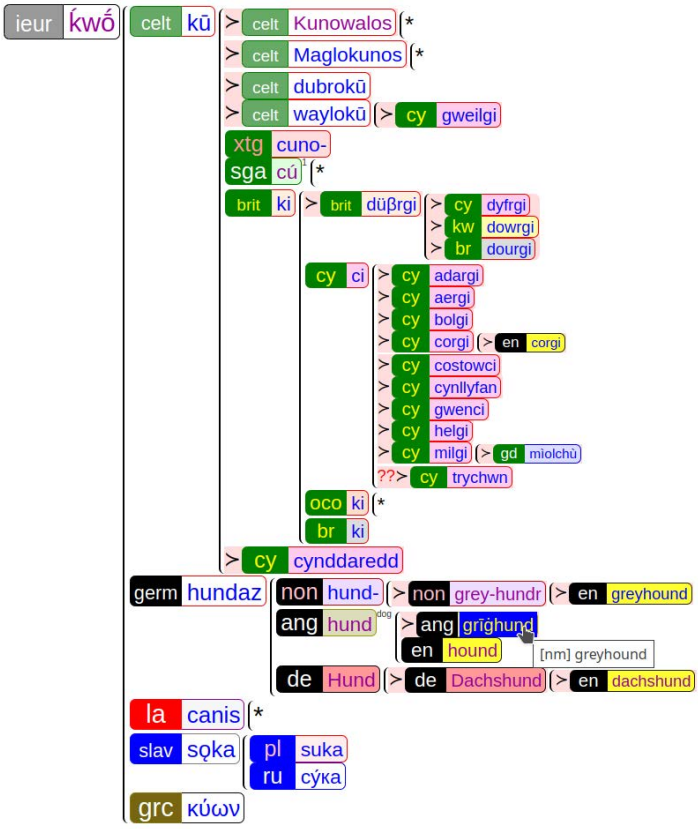


Figure 3: Tree with "Show super-elements" turned on.

The word cy:*trychchwn* is preceded by two red question marks, and hovering over them will display "probability 0.6". It is sometimes uncertain as to whether words are related, and Bunadas can take this uncertainty into account by attaching a probability to whether a word rightly belongs to a cluster. When generating trees, Bunadas naively multiplies any probabilities it encounters, and if the resulting probability drops to 0.5 or lower, it stops there and goes no further along that particular path. So, for example if there is a probability 0.6 that word A is an ancestor of word B, and a probability 0.7 that word B is an ancestor of word C, then Bunadas naively assumes that there is a probability $0.6 \times 0.7 = 0.42$ that word A is related to word C, and therefore it does not (unless there is other evidence) show word C or its descendants in the tree it generates starting from word A, because the connection is too dubious.

The word celt:*Kunowalos* has a (* after it, as do four other words. This is because I have "pruned" the tree at these points, compressing branches to reduce clutter and fit the tree into a reasonably sized figure. Any branch can be pruned by clicking the language tag of the headword and can be "unpruned" by clicking again or by clicking the * symbol. The pruning is "remembered" in the URL, which is very useful when passing Bunadas tree URLs to other people as illustrations of particular points.

As well as the option "Show super-elements" in Bunadas trees, there is an option "Show sub-elements". But here Bunadas has to be very careful. If both options are switched on, it will happily go down to sub-elements when constructing trees and then back up again: from cy:*corgi* to cy:*ci* to cy:*milgi* for example, which is fine because the words are all related. However, the Bunadas algorithm will not ever go up and then back down. Otherwise, it would go from de:*Hund* to de:*Dachshund* to de:*Dachs* and claim wrongly that de:*Hund* and de:*Dachs* were related.

In addition to the ≻ symbol, Bunadas has two other symbols which are important for relating words. The ≫ symbol shows that one "word" (or prefix or suffix) is only a very minor component of another word. Conversely the ⪰ symbol is for showing that two words are substantially the same; that one is a very major component of the other. Thus the ⪰ symbol is typically used when linking a word with its root. The ≫ symbol is used when linking words to minor morphemes such as prefixes and suffixes. In analyzing the English word en:*only*, for example, it could be linked to the suffix en:*-ly* using the ≫ symbol and linked to the English word en:*one* using the ⪰ symbol. When constructing trees, the rule "no going up then back down", which applies to the ≻ symbol, does *not* apply to the ⪰ symbol. When constructing trees, Bunadas does not normally follow ≫ links at all – otherwise it would construct many false links via trivial prefixes and suffixes. The ≫ symbol can also be used to record that a word is a calque of a word in another language. Thus, the Old Irish *día lúain* 'Monday', for example, is derived from the native words *día* and *lúan*, and the relationship is shown with the ≻ symbol. It is a calque of Latin *diēs Lūnae*, and this connection can be shown with the ≫ symbol, ensuring that it is not followed when constructing trees.

The ≻ symbol is most often used when words have combined into one by compounding, as in English *seafarer*, German *Seefahrer*, or Proto-Celtic *tegoslougom*. However, it can also be used where two words which were near homonyms have merged into one. Another situation in which the ≻ symbol may be used is where a word may have come partially via one route, partially via another – such as where an Old Irish word may have come partially from an Old Norse word but partially from a similar Old English word, and both of these have in any case

come from the same Proto-Germanic word. In this case, it is possible to get the best of both worlds by linking the Old Irish word using ≻ to the Old Norse and Old English words, but also linking it directly to the Proto-Germanic ancestor, maybe setting the distance metric here as 2. Alternately, we might want to give it probability 0.5 as having come from Old Norse, 0.5 from Old English, while linking it with certainty to the Proto-Germanic ancestor.

Most of the time it is obvious which symbol is appropriate when linking words together. Sometimes, though, it is not certain whether it is best to break a word down into two equal-status components using ≻ for each; or whether to apply ⪰ to one and ≫ to the other, and if so which of them should be considered the major component (which will therefore appear in trees) and which the minor. It is important for anyone editing Bunadas to maintain consistency, and if say, an Old French word is broken down in a particular way, then the Modern French word should be broken down in the same way.

The idea of a tree of descendants can go "wrong", or at least exhibit somewhat pathological behavior. Take, for example, the tree of descendants for PIE *tḱey-*. In all the descendants, we can at least see some reflex of the *tḱ* from the PIE root – except in the branch derived from Latin *sinō* and *pōnō*:

(1) *tḱey* → ieur:*tḱi-né-ti* → la:*sinō* → la:*pōnō* → la:*conpōnō* → en:*compose*

The *s* of Latin *sinō* comes from the *tḱ* of *tḱey-*, but this has disappeared in *pōnō* after prefixing by *po-*, and *pōnō* has absolutely nothing left of the original *tḱey-*; everything it contains comes from prefixes and suffixes added later. This is even more extreme by the time we reach English *compose*. On each step of the way from *tḱey-* to *compose*, each word has clearly been derived from the word immediately before it, and yet no smidgeon of *tḱey-* can be found in the word *compose*. Other examples can be found, such as PIE $g^{h}ed$- → English *prison* (as well as the more transparent English *get*). Fortunately, this situation does not occur often enough to be a big problem, and the trees generated by Bunadas are nearly always informative and meaningful.

There is a question as to whether we need the concept of "cluster" at all, or whether it would be better to relate words directly to one another, going word-word rather than word-cluster-word. There are advantages and disadvantages to each approach. From a purely theoretical point of view, both approaches are equivalent and can be converted into one another. Word-word connections are obviously a special case of small clusters. And clusters can be broken down automatically into word-word connections, albeit rather a lot of them in the case of a large cluster having no "center" word. Having direct word-word connections

would simplify the programming a bit, perhaps simplify the logic a bit, and would speed up the computations slightly. The clusters, though, are very useful to human editors, and facilitate meaningful groupings which are easier to assess and to manage. The Bunadas editing facilities allow words to be drag-and-dropped from cluster to cluster, which is very useful.

The Bunadas editing tools can be tried out by anyone who is interested, simply by swapping to the bunTest database in the dropdown on the main Bunadas page. The Bunadas programs are all publicly available on GitHub[9] to anyone who is interested. The /cloning webpage[10] gives details, and even includes a moderately up-to-date dump of the entire Bunadas database. The Bunadas user interface is available in a dozen languages so far, and there is a good translation interface so it would be easy to add still more interface languages. As can be seen from GitHub, the programs are written in PHP, with MariaDB (or MySQL) as the backend relational database, and with JavaScript/AJAX used to help provide a responsive user interface.

3 Orthography, and which wordform to use for Bunadas words

It is important for Bunadas to be consistent about orthography so that the same word is not added to Bunadas multiple times, to the confusion of users. Bunadas nearly always uses the orthographic guidelines defined by Wiktionary for the various languages.[11] This is particularly important for the reconstructed proto languages, where no historic spelling system exists, and diverse systems have been used by different authorities and authors. Wiktionary has created detailed orthographic guidelines for Proto-Indo-European, Proto-Celtic, Proto-Brythonic and , among others. For languages where Wiktionary sometimes uses more detailed orthography, sometimes less detailed, Bunadas takes care to use the more detailed orthography. This is the case for Ancient Greek, for example, where the particular entries have detailed diacritics marking breathings, short vowels, etc., but the internal links in Wiktionary may use less detail.

At present Old Irish is an exception to the rule that Bunadas follows Wiktionary spelling guidelines. Here Bunadas generally follows the spelling in DIL, the Dictionary of the Irish Language. This is because DIL is so authoritative

[9] https://github.com/caoimhinsmo/bunadas
[10] https://www.smo.uhi.ac.uk/teanga/bunadas/cloning/
[11] https://en.wiktionary.org/wiki/Category:Wiktionary_language_considerations

and so readily available online, but most of all because of the huge contribution to Bunadas derived from the *Innéacs Nua-Ghaeilge don DIL* (De Bhaldraithe 1981) about 13,000 Old-Irish–Modern-Irish pairs. In particular, this means that Bunadas currently uses the hyphen (-) rather than the raised dot (·) to show the "joint" of the verbal complex. It is expected that Bunadas will move at some stage to using the Wiktionary orthography for Old Irish. This will not, though, cause the link with DIL to be lost, because each Old Irish word in Bunadas is already linked to its numeric entry number in DIL.

Just as important as consistency in orthography, is consistency in which word-form to use for Bunadas words – whether, for example, to use the first-person-singular for verbs, or else the third person singular, or singular imperative, or infinitive. Here again, Bunadas follows the Wiktionary preference for each particular language: the first singular for Latin verbs, for example, but the infinitive for French and Old French verbs. Wiktionary does in fact also try to include brief entries for the infinitive of Latin verbs, first singular of French verbs, etc., and indeed for nearly all other grammatical forms, but its preferred form is clear for each language. Bunadas goes for much less morphological granularity than Wiktionary does, so as to avoid clutter which would be unhelpful to the user. To illustrate from English, Bunadas includes both the verb *state* and the noun *statement*, but it does not include verbforms *stating*, *stated*, and *states*, whereas Wiktionary includes all of these. This difference in morphological granularity means that while Wiktionary can link a French verb (in the infinitive) to the Latin infinitive from which it is derived, Bunadas links it to the Latin first person singular; and while Wiktionary can link a French noun to the Latin accusative from which it is derived, Bunadas links it to the Latin noun in the nominative.

In the case of Modern Irish, the need to standardize to the singular-imperative of verbs required a major conversion exercise. The singular-imperative is used by Wikipedia and by most dictionaries, but unfortunately the third singular is the form which was used by the *Innéacs Nua-Ghaeilge*, so all of these had to be converted, and disambiguators added where required.

In the case of suppletive verbs or suppletive nouns, the idea of sticking to a standard word form for each language breaks down. The English past tense *went*, for example, cannot be simply covered by the verb *go* as in many dictionaries – it requires a separate entry in Bunadas. The choice of which word form to use to represent a group of suppletive forms often has to be made on an ad hoc basis.

4 What Bunadas does and does not do

What Bunadas primarily does is exploit connectivity. If word A is etymologically related to word B, and word B is etymologically related to word C, then word A is etymologically related to word C, and so on. There is no need to store separately the information that A is etymologically related to C. This avoids duplication and a lot of the gaps in conventional etymological dictionaries.

What Bunadas does not do is concern itself with how or why or even in what direction A is etymologically related to B: whether A is inherited from B, or is a modern borrowing of B, or vice versa, or whether they are both derived separately from a common ancestor. Bunadas can link A with B in all these different circumstances. Of course, Bunadas can give clues about how A might be related to B. If A is a Latin word at the center (metric distance 0) of a cluster which also contains B and other modern Romance words, this is a very strong clue that B is derived from A rather than the other way round. If a French word B is linked straight to a Latin word A, while another French word D is linked via Middle French and Old French words to A, then this indicates strongly that French word D is inherited, while B may or may not be a modern borrowing. As more words in intermediate languages are added to Bunadas, these clues become stronger, but they are still just clues. Bunadas has a free-text comments field for each word where information about derivation can be added. All these things, though, are peripheral to Bunadas. They are not part of its core functionality, which is about connectivity. Bunadas contains no information at all about sound laws: about how an inherited Modern French word D, for example, might have been shaped by going through the mangle of Middle French and Old French and Vulgar Latin sound changes. It is conceivable that sound law information might somehow someday be added to Bunadas, but this is not planned or envisaged in the near future.

5 The future

Where does Bunadas go from here? Obviously adding more words in more languages would increase its functionality. Adding more words would make the trees more cluttered, but this could be alleviated by improving the filtering/pruning facilities: by introducing filtering by language family for example.

Bunadas was initially populated by about 5000 words from the *Stòr-fhaclan Co-dhàimheil Ceilteach* (Celtic Cognates Database), a project which had become moribund because its rectangular structure, languages × word groups, was too

inflexible and made adding data too difficult. To this were added about 12,000 Old-Irish–Modern-Irish word pairs from the *Innéacs Nua-Ghaeilge don DIL* which Kevin Scannell has put online as *Droichead DIL*.[12] Bunadas also contains about 21,000 Manx words and word forms and cognates (mostly Irish), a donation from Kevin Scannell, although these have not yet been fully integrated into the system. All the other thousands of words have been added laboriously by hand, using resources such as GPC for Welsh (Thomas et al. 2022) and DIL for Irish (Toner et al. 2019), but most of them, especially for the non-Celtic languages, are copied from information in Wiktionary. This work continues, even though Bunadas has never been a funded project.

The sources of the etymological information in Wiktionary are documented in the entries in Wiktionary itself. There are hundreds of them, both books and papers, far too many to list here, but two which could be mentioned for the Celtic languages are Matasović (2009) and Zair (2002). Apart from Wiktionary, the Celtic language information in Bunadas is gleaned from dozens of sources, such as GPC and DIL and all the usual Celtic language etymological dictionaries, including Deshayes (2003), Henry (1900), and MacBain (1911).

100,000 words is a good start but is a drop in the ocean compared to all the words in all the languages of the world. By adding more words to it, Bunadas could be transformed into a useful resource for not just the Celtic languages, but lots of other languages too. An obvious question is, if much of the data is being copied over from Wiktionary, would it not be better to scrape data automatically from Wiktionary, since this is all in the public domain? This thought has already occurred to various very knowledgeable data scientists, who have constructed facilities based on etymological data scraped from Wiktionary.[13] The results so far are useful, have more words and more languages than Bunadas has, but they have nothing to compare with the impressive, broad trees constructed by Bunadas. The problem is that even though the etymological information contained in Wiktionary is generally very good and is improving by the week, the way it is structured is dire. Wiktionary does not give unique IDs to words, so links between words are based on spelling and can be broken if the spelling is changed. Wiktionary does not distinguish properly between homographs. It piles all the homographs, both between languages and within languages, onto the one webpage. Wiktionary currently says, for example, that English *bole* 'tree trunk' is from Norse *bolr*, and that *bolr* is from PIE *b^{h}el-*, but it does not clarify which of the four different PIE roots *b^{h}el-* it means: the link just leads to the four of them.

[12] https://cadhan.com/droichead/

[13] Lexvo: http://www.lexvo.com/ (De Melo 2015); Rabbitique: https://www.rabbitique.com; Etymologeek → Cooljugator: https://cooljugator.com/etymology

I believe that the long-term solution lies in Wikidata Lexical Data.[14] The various Wikipedias, especially for smaller languages, are increasingly starting to automatically incorporate data from Wikidata, which is a common database of information of all kinds, and Wikidata Lexical Data is its extension into lexicography. The words in Wikidata Lexical Data do have proper unique identifiers, 'L' numbers (such as L6419 for en:*hound*), corresponding to Wikidata's 'Q' numbers. The long term aim, I believe, should be for etymological data to be shared in Wikidata Lexical Data, from where it can be accessed and used by all the various Wiktionaries, http://ru.wiktionary.org, http://fr.wiktionary.org, etc., as well as http://en.wiktionary.org, and used also by Bunadas and other such facilities. Wikidata Lexical Data, as it stands at present, has a way to go before it is ready for this. It has not yet fully embraced the essential idea of separating out etymologically distinct homographs as distinct words. It lacks the categorization of types of word components which Bunadas expresses using the $\succ$, $\succcurlyeq$, and $\gg$ symbols. It is likely that it will be several years before Wikidata Lexical Data is ready to provide a foundation for the etymological information expressed in Bunadas and in the various Wiktionaries. So, in the meantime Bunadas can continue on its present course, providing useful facilities especially for the Celtic languages, and exploring what will be possible and useful in an etymological facility for all languages once the data foundation is there.

References

Banksira, Degif Petros. 2000. *Sound mutations: The morphophonology of Chaha*. Amsterdam: John Benjamins.

Bell, Elise, Diana B. Archangeli, Skye J. Anderson, Michael Hammond, Peredur Webb-Davies & Heddwen Brooks. 2023. Northern Welsh. *Journal of the International Phonetic Association* 53(2). 487–510. DOI: 10.1017/s0025100321000165.

Bosch, Anna & James Scobbie. 2009. Fine-grained morpho-phonological variation in Scottish Gaelic: Evidence from the Linguistic Survey of Scotland. In James N. Stanford & Dennis R. Preston (eds.), *Variation in indigenous minority languages* (IMPACT: Studies in Language and Society 25), 347–368. Amsterdam: John Benjamins.

De Bhaldraithe, Tomás. 1981. *Innéacs Nua-Ghaeilge don dictionary of the Irish language*. Baile Átha Cliath (Dublin): Acadamh Ríoga na hÉireann (Royal Irish Academy).

[14] https://ordia.toolforge.org

De Melo, Gerard. 2015. Language-related information for the linguistic linked data cloud. *Semantic Web* 6(4). 393–400. DOI: 10.3233/SW-150171.

Deshayes, Albert. 2003. *Dictionnaire étymologique du Breton*. Douarnenez: Le Chasse-Marée.

Henry, Victor. 1900. *Lexique étymologique des termes les plus usuels du Breton moderne* (Bibliothèque Bretonne Armoricaine; III). Rennes: J. Plihon et L. Hervé.

MacBain, Alexander. 1911. *An etymological dictionary of the Gaelic language*. Stirling: Eneas Mackay.

Matasović, Ranko. 2009. *Etymological dictionary of proto-celtic*, vol. 9 (Leiden Indo-European Etymological Dictionary Series). Leiden: Brill.

Ó Donnaíle, Caoimhín P. 2014. Tools facilitating better use of online dictionaries: Technical aspects of multidict, wordlink and clilstore. In *Proceedings of the first Celtic language technology workshop*. https://aclanthology.org/W14-4603.pdf (5 June, 2025).

Sims-Williams, Patrick. 1990. Dating the transition to Neo-Brittonic: Phonology and history, 400–600. In Alfred Bammesberger & Alfred Wollmann (eds.), *Britain 400–600: Language and history*, 217–262. Heidelberg: Carl Winter Universitätsverlag.

Thomas, R. J., Gareth Bevan, Patrick Donovan & Andrew Hawke. 2022. *Geiriadur prifysgol cymru online*. https://www.welsh-dictionary.ac.uk/ (5 June, 2025).

Toner, Gregory, Máire Ní Mhaonaigh, Sharon Arbuthnot, Dagmar Wodko & Maire-Luise Theuerkauf. 2019. *eDIL: electronic dictionary of the Irish language. based on the contributions to a dictionary of the Irish langauge (Dublin: Royal Irish Academy 1918-1976)*. https://dil.ie (5 June, 2025).

Zair, Nicholas. 2002. *The reflexes of the Proto-Indo-European laryngeals in Celtic*. Leiden: Brill.

Chapter 11

Laryngeal contrast in the Celtic languages: Variation, typology, and history

Pavel Iosad
University of Edinburgh

This chapter addresses the nature of laryngeal contrast throughout the history of the Celtic languages. In the present day, there is significant phonetic and phonological variation across the languages in distinctions among two series of stops, while the historical data is ambiguous in multiple ways. The chapter re-evaluates the issues in light of some recent findings in phonetic and diachronic typology. I argue that the "aspirating" system of many present-day languages (contrasting aspirated /pʰ tʰ kʰ/ stops with a variably voiced /b d g/ series) most likely arose around the mid 1st millennium CE "syncope period" in both Gaelic and Brythonic. I show there are significant difficulties in projecting it towards the earlier period, where a "voicing-based" system (unaspirated /p t k/ vs. categorically voiced /b d g/) is more compatible with the observed sound changes. This analysis implies that the rise of the aspirating system aligned with a suite of drastic phonological changes affecting the Insular Celtic languages in the sub-Roman period, and underscores the importance of synchronic variation for reconstructing the early history of the Celtic laryngeal contrast.

1 Introduction

The aim of this chapter is to consider the history of laryngeal contrast in the Celtic languages, especially the phonetics and phonology of stops. Most present-day Celtic languages distinguish between two series of stops, spelled in most

Pavel Iosad. 2025. Laryngeal contrast in the Celtic languages: Variation, typology, and history. In Andrew Carnie, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond & Luis Irizarry (eds.), *Foundational approaches to Celtic linguistics*, 383–415. Berlin: Language Science Press. DOI: 10.5281/zenodo.15654877

contexts as ⟨p t c/k⟩ and ⟨b d g⟩.[1] Their phonetic realisations vary non-trivially across the family: to avoid ambiguity, I will use the terms FORTIS and LENIS respectively to refer to the two series, eschewing the terminology of "voiced", "voiceless", or "(un)aspirated" to avoid confusion between phonological and phonetic description.[2] Synchronically, the present-day Celtic languages are fairly similar in the (morpho)phonological patterning of stops, but are considerably more diverse in how the contrast is phonetically implemented. Diachronically, the trajectories of sound change that led to the present-day states of the languages differ more significantly, and, as I will show in this chapter, this provides us with a wealth of information for phonetic analysis and typological comparison.

In order to consider the history of laryngeal contrast, in Section 2 I sketch the patterns observed in the present-day languages. Section 3 lays out the arguments made in favour of a view of earlier stages of Celtic as an "aspirating" language characterized by aspiration of the FORTIS series and variable voicing of the LENIS series of stops; in various guises, this analysis has long been commonly accepted as correct. In Section 4 I lay out the relevant sound changes and consider whether they are phonetically and typologically plausible from an "aspirating" starting point. Section 5 synthesizes the findings and lays out the key argument of this chapter that "aspirating" systems are a relatively recent innovation in the history of the attested Insular Celtic languages, possibly dating to no earlier than the mid 1st millennium CE. Section 6 concludes the chapter with a range of further considerations, with specific attention to reconstructing variation in the proto-language and its consequences particularly in the case of Celtic.

[1] I use ⟨angled brackets⟩ to refer to orthographic representations, ALL CAPS for abstract phonological categories, |pipes| to signal phonological feature specifications, /slashes/ for phonological underlying specifications, and [square brackets] for phonological surface representations.

[2] The terms "fortis" and "lenis" themselves, of course, do also have phonetic connotations and have been used to refer to aspects of phonetic realisations: for a review, see Honeybone (2008). It would have been preferable to use even less suggestive terminology: one possibility is calling the ⟨p t k⟩ series *tenues* and ⟨b d g⟩ *mediae*, a usage formerly common especially in Indo-European (including Celtic) comparative philology. However, this terminology is now so unfamiliar that I have opted for "fortis" and "lenis".

2 Laryngeal phonology in present day Celtic

2.1 Theoretical preliminaries

The analysis of laryngeal stop contrasts in this chapter is informed by the influential framework of Laryngeal Realism (a term due to Honeybone 2005). It was formulated as an alternative to a view common in mainstream generative phonology, in which two-term systems contrasting ⟨p t k⟩ and ⟨b d g⟩ series were generally analysed as specified [−voice] and [+voice], respectively. Under this "broad interpretation of the feature [±voice]" (Hall 2001), many cross-linguistic differences in the phonetic correlates of the two specifications are considered irrelevant, as long as voicing participates in the realisation of the contrast in some way. Therefore, languages like French [stan1290] and Ukrainian [ukra1253], in which FORTIS stops lack significant aspiration and LENIS stops have categorical voicing in all contexts could be analysed as identical to languages like English [stan1293] and German [stan1295] with aspiration of the FORTIS stops in many contexts and variable presence of voicing in the LENIS series: aspiration in the latter can be either ignored or considered redundant. It was only languages like Icelandic [icel1247], which distinguish between [pʰ tʰ kʰ] and [p t k], that needed a different analysis, since both series are [−voice]: in this case, with a [+spread glottis] specification for the FORTIS stops.

For proponents of Laryngeal Realism, the analysis of such systems should take into account both the phonetic properties of the contrast and the stops' morphophonological behaviour. Featural specifications are required to be supported by evidence of two kinds. First, the relevant feature's phonetic correlate (such as voicing or aspiration) should consistently participate in the realisation of the stop series. Second, one should look for (morpho)phonological evidence that the specification is active. For instance, if a stop bears the |voice| specification, then this feature can trigger voicing assimilation, or be restricted from appearing in "weak" positions, such as word- or syllable-finally, resulting in sound patterns such as word-final devoicing. Widespread adoption of versions of Laryngeal Realism has been encouraged by numerous cases where the two sets of criteria – phonetic realisation and phonological activity – point in the same direction. In this way, languages can be classified as belonging to one of a small number of "types". Thus, English and German have been shown to combine aspiration of FORTIS stops with their phonological visibility on the one hand, and an absence of phonetic or phonological activity of voicing in the LENIS series; conversely, languages like Catalan [stan1289] or Bulgarian [bulg1262] combine categorical voicing of LENIS stops with evidence for phonological activity of this series, while FORTIS stops are both unaspirated and relatively inert phonologically.

This happy confluence has led to the emergence of a particularly strong version of Laryngeal Realism, in which phonetic and phonological criteria must always point in the same direction. Conceptually, this prediction arises because no mechanism within the theory allows for such conflict. Under this view, demonstrating that a particular phonetic cue is obligatorily associated with some category of segments is sufficient to infer the presence of the corresponding featural specification, even in the absence of clear (morpho)phonological evidence (e. g. Beckman et al. 2013). The converse corollary also holds: (morpho)phonological evidence alone can be used to determine a language's place in the typological space. This is particularly useful for historical phonology. While phonetic detail is usually not directly recoverable, one can be on surer ground with respect to morphophonological patterns based on philological or dialectal evidence. If the corollary is correct, one is in a position to make inferences about phonetic detail in the past based on phonological reconstruction. For example, Spaargaren (2009) argues that patterns of laryngeal assimilation in Old English (as in *cēp-**te*** 'keep.PST.3SG' vs. *freme-**de*** 'do.PST.3SG', *hīer-**de*** 'hear.PST.3SG') show greater phonological activity of the |fortis| series of stops, which should lead us to reconstruct Old English |fortis| stops as being aspirated.

This very strong version of Laryngeal Realism has been challenged on a number of grounds. Salmons (2020) argues that a phonetic specification does not necessarily imply the presence of the relevant phonological feature, as in those Germanic languages like Dutch [dutc1256] and perhaps Central Swedish [stan1279] where |lenis| stops are consistently voiced but remain relatively inert phonologically. Kirby & Ladd (2019) question the privileged role of aspiration (voice onset time) in the typology, pointing to the important role of other cues. Most controversially, Cyran (2013) and Iosad (2017) develop case studies to show that phonetic and phonological criteria can be in outright conflict (for a rejoinder, however, see Raimy 2021).

In this chapter, I will use Laryngeal Realism terminology where appropriate to refer to covarying aspects of phonetic and phonological behaviour. With the exception of Breton and possibly several varieties of Scottish Gaelic, the present-day Celtic languages belong to the "aspirating" type, as I will now discuss in more detail.

2.2 The Gaelic languages

In Laryngeal Realism terms, the Gaelic languages clearly belong to the "aspirating" type, with marked |spread glottis| stops in the ⟨p t k⟩ series and unspecified ⟨b d g⟩ stops.

There is near-universal agreement across auditory and instrumental descriptions that the FORTIS stops are postaspirated in both Irish (e.g. Ní Chasaide 1985, Ó Raghallaigh 2014) and the majority of Scottish Gaelic varieties (e.g. Ní Chasaide 1985, Nance & Ó Maolalaigh 2021) in the onset of a stressed syllable. Furthermore, FORTIS stops are preaspirated in postvocalic (or perhaps more precisely postsonorant) position in much of Scottish Gaelic, and likely also at least in some varieties of Irish (Iosad 2025).

Phonologically, |spread glottis| is active in the FORTIS series. The clearest evidence comes from Irish, which has a pattern of coalescence of a LENIS obstruent with a following /h/ that results in a FORTIS consonant. It is observed, for instance, with the future/conditional verbal suffix spelled ⟨-f⟩: Modern Irish (Ir*lig* 'let.IMP.2SG' with [gʲ], *lig-f-idh* 'let-FUT-3SG' with [kʲ] (Sommerfelt 1960, Ó Buaachalla 1985). Similar behaviour is seen with the passive participle suffix *-tha*: *sníomh* 'spin.IMP.2SG' with [v]/[β], *sníofa* 'spun'.)

Evidence also comes from initial mutation: in lenition, FORTIS stops /pʰ tʰ kʰ/ (and their slender counterparts) alternate with the voiceless fricatives [f h x], and it is widely accepted that such fricatives are phonologically |spread glottis| ("Vaux's Law", after Vaux [1998]). Assuming that lenition alters only the manner specification of the stops (*modulo* any necessary place changes), this can be taken as further evidence that the FORTIS stops are |spread glottis|.

The situation in Scottish Gaelic is somewhat less conclusive: due to morphological differences, it does not show as many clear examples of the assimilation pattern.[3] It maintains the pattern of mutation, but without clear language-internal evidence for |spread glottis| activity in the fricatives, the argument is much weaker. Nevertheless, there is also no clear evidence against an "aspirating" analysis, and the phonetic data clearly points in the "aspirating" direction just like in Irish.

With regard to the LENIS stops, one is slightly hampered by the lack of detailed instrumental studies, especially of Irish; however, see Ní Chasaide (1985) on Ulster (Gaoth Dobhair). The consensus is that Irish LENIS stops are well-behaved for an "aspirating" language, with variable voicing especially after sonorants (cf. again Ó Raghallaigh 2014). Traditionally, the presence of voicing in the LENIS series is considered an isogloss separating Irish from Scottish Gaelic, which is generally described as having voiceless unaspirated ("short-lag" VOT) LENIS stops (O'Rahilly 1936: 146 sqq.; Ó Murchú 1985, Robinson & Ó Maolalaigh

[3]The future and conditional are formed without the *-f* element in Scottish Gaelic, and the passive participle is not generally productive. The devoicing pattern can be observed in head-final compounds, but it is not obvious that this is a productive rule, either (cf. Stifter 2015).

2007). This is largely confirmed by instrumental studies such as Ladefoged et al. (1998), although some voicing is possible even here (Nance & Stuart-Smith 2013); Ní Chasaide (1985) also reports (albeit with small sample sizes) some dialectal variation across different parts of the Western Isles.

The literature on Laryngeal Realism does allow for "aspirating" languages to show either variably voiced or fully voiceless LENIS stops. Irish patterns with languages like English and German, whilst Scottish Gaelic is similar to languages like Danish [dani1285] and Icelandic, which are all considered to be "aspirating" by Beckman et al. (2013). Both Irish and Scottish Gaelic also demonstrate a pattern common in Germanic "aspirating" languages whereby the contrast between the two series is neutralized after a tautosyllabic fricative (most notably in the clusters *sp st sk*), with a voiceless unaspirated stop as the outcome. The usual interpretation of such structures (e. g. Salmons 2020) is that the stop and the fricative share a |spread glottis| specification. Thus, from the perspective of Laryngeal Realism both Irish and Scottish Gaelic present as "aspirating" languages.

2.3 The Brythonic languages

The phonetics and phonology of Welsh are likewise straightforwardly compatible with an "aspirating" analysis. The FORTIS stops are postaspirated before a stressed vowel, and at least variably preaspirated after a vowel (Ball & Williams 2001, Jones 1984, Morris & Hejná 2020, Iosad 2023). The LENIS stops are generally described as variably voiced, certainly in South Welsh (Jones 1984); there is a perhaps somewhat surprising dearth of instrumental investigations of North Welsh, but Bell et al. (2023) report a Scottish Gaelic-like system of categorically voiceless LENIS stops. As in the Gaelic languages, the contrast is neutralized to a voiceless unaspirated stop after a tautosyllabic FORTIS fricative.

Phonologically, Welsh is a very clear example of a language showing activity of |spread glottis| in the FORTIS series and inertness of the LENIS stops: for example, as in Irish, a suffix-initial /h/ coalesces with LENIS stops, as in [cy] *gwag* 'empty', *gwacáu* 'to empty'. This evidence is considered in detail by Iosad (2023), to which the reader is referred for further discussion. Thus, Welsh is quite unproblematic as an "aspirating" language.

The phonetic detail of Cornish is not available, but the morphophonological patterns it shows are quite parallel to those seen in Welsh, notably in devoicing ("provection") patterns that are (residually) attested in the comparative forms of adjectives ([mco] *teg* 'fair', *tecka* 'fair.CMP', cf. *gwan-ha* 'weak-CMP') and in the subjunctive (*[d]eppr-o* 'eat.SBJV-3SG' to *dybry* 'eat'); cf. Williams (2011).

2.4 Voicing in Breton and Scottish Gaelic

So far, there has been no discussion of any Celtic variety that demonstrates categorically voiced LENIS stops. There are two groups of present-day languages where such representation is found.

The first of these is Breton. As discussed in Iosad (forthcoming), phonetically it is much more reminiscent of a "voicing" language similar to its close neighbour French, with fully voiced LENIS stops and a voiceless unaspirated FORTIS series. However, its morphophonology is more like that of an "aspirating" language. In particular, just like Irish, Welsh, and Cornish, Breton shows a pattern of LENIS stops coalescing with suffix-initial /h/ to their FORTIS counterparts: [br]*gleb* 'wet',[4] *glep-oc'h* 'wet-CMP' patterns just like W *gwlyb*, *gwlypach*. Even patterns superficially resembling those common in "voicing" languages, such as final devoicing, turn out on closer inspection to involve greater markedness of the FORTIS series, meaning that the phonology of Breton is more like that of an "aspirating" language just like its closest relatives. See Iosad (2017) and Iosad (forthcoming).

Consistent closure voicing is also found in some Scottish Gaelic varieties. First, Henderson (1903) reports that all LENIS stops are voiced in "peripheral" dialects, such as Sutherland and South Argyll. Allen (2022) examined the published dialect survey (Ó Dochartaigh 1994) and found consistent voicing in the LENIS series in parts of eastern and south-eastern Sutherland (points 141–147), and variable presence of voicing in north and west Sutherland and Caithness (points 128–140).

Second, closure voicing in stops is associated with the process of post-nasal sandhi ("eclipsis") across many dialects of Scottish Gaelic (Ó Maolalaigh 1995, Bosch & Scobbie 2009). Both FORTIS and LENIS stops participate in various progressive assimilations following nasals, either word-internally or (a common source of alternations) after nasal-final clitics. There is significant variability in the precise outcomes of these, as Bosch & Scobbie (2009) document, but closure voicing of the stop is one possible result.

The two stop series tend to remain distinct (for instance, as breathy voiced vs. plain voiced stops) in these contexts: Gairloch (Wentworth 2005) *an doras* 'the door' [dɔrəs] vs. *an toll* 'the hole' [d̤o͡ul̪ˠ] (unmutated [tɔrəs], [tʰo͡ul̪ˠ]).[5] Subse-

[4]The form *gleb* is realised in isolation with a voiceless stop due to final devoicing, but cf. *glebiañ* 'to wet'.

[5]There is in many, probably most, dialects a strong tendency to show different developments in "close sandhi" (after proclitics and in close compounds like *lon-dubh* 'blackbird') and word-internally (as in *cinnteach* 'certain'): the series tend to remain distinct in the former contexts but often show a tendency towards merger in the latter, with most stops becoming LENIS and realised with voicing. The matter requires much further scrutiny.

quent loss of the nasal may lead to outcomes that can be interpreted in terms of a third distinct phonemic series characterized by a |voice| specification; this is proposed, for example, for Applecross by Ternes (2006) and Easter Ross by Watson (2022).

Because of this, we can only rarely analyse postnasal voicing in terms of a "voicing" system: postnasal processes do not generally result in a system contrasting a voiceless (let alone unaspirated) FORTIS series /p(ʰ) t(ʰ) k(ʰ)/ and a consistently voiced LENIS /b d g/. The only exception here, as Bosch & Scobbie (2009) note, are peripheral dialects of the Sutherland type, but within the broader Celtic context this phenomenon is quite marginal (in more than one sense).

To summarize this section, most present-day Celtic languages are well-behaved representatives of the "aspirating" type under Laryngeal Realism. There is ample phonetic and phonological evidence in favour of associating the FORTIS stops with a featural specification such as |spread glottis|, and very little to justify assigning a feature such as |voice| to LENIS stops. Even where consistent voicing does occur in the LENIS series, it does not clearly justify a reclassification of the variety in question as representing the "voicing" type. In the next section, the prevailing view of laryngeal contrast in an earlier era is considered.

3 The aspirating view of Proto-Celtic

Unsurprisingly given the situation discussed in Section 2, Proto-Celtic is usually reconstructed with two stop series: a FORTIS category corresponding to "traditional" Proto-Indo-European "voiceless" **p t k kʷ*, with a very early (but not Proto-Celtic) change of **p* to zero, likely via a labial fricative and then **h* in most contexts. The LENIS series **b d g gʷ* represents traditional Indo-European "voiced" and "voiced aspirate" stops. These stops could be either short or long (geminate). While FORTIS geminates were not at all unknown, LENIS geminate stops were rather more marginal, occurring primarily as secondary developments of other clusters or across a morpheme boundary; for the details of gemination in Celtic, see Stifter (2024).

A very simple interpretation is to treat the FORTIS series as (plain) "voiceless" and the LENIS series as "voiced". However, direct evidence in favour of this account is meagre. Much of the Continental Celtic inscriptional material is not probative at all: it is preserved using writing systems such as the Iberian and Lugano alphabets that generally do not distinguish laryngeal contrast in stops. Celtic material in the Greek and Latin scripts generally uses orthographic ⟨p t c⟩ *resp.* ⟨π τ κ⟩ for the FORTIS stops and ⟨b d g⟩ *resp.* ⟨β δ γ⟩ for their LENIS counterparts. However, there is some potentially useful graphical hesitation. The Gaulish

material in the Latin alphabet presents non-trivial amounts of vacillation: especially frequent is ⟨c⟩ for expected [g], and we also find some confusion between ⟨t⟩ and ⟨d⟩. Furthermore, a few Greek-script inscriptions witness ⟨θ⟩ and ⟨χ⟩ for the FORTIS stops. Following Gray (1944) and Watkins (1955), this vacillation is often taken to indicate that singleton stops were pronounced differently in Gaulish and in Latin. A common view is that Gaulish /b d g/ possessed some "fortis" property that their Latin counterparts lacked, precluding orthographic ⟨b d g⟩ from representing these LENIS stops: one possibility is that they would not be consistently voiced, being therefore liable to be identified with Latin or Greek voiceless stops.

Given the "aspirating" systems of the present-day languages, it is also possible to simply interpret Proto-Celtic as long-lag VOT ("aspirated") stops in the FORTIS series. Already, Pedersen (1909) and Lewis & Pedersen (1937) projected the present-day aspiration into Proto-Celtic. In a series of recent publications, Eska (2017; 2018; 2019; 2020) forcefully restated this position within the theoretical framework of Laryngeal Realism, arguing that Proto-Celtic is best analysed as an "aspirating" language, and thus that the present-day systems continue the original pattern unchanged. He adduces evidence of two kinds for this proposition.

In terms of direct evidence, his proposal is supported by the interpretation of the Gaulish material discussed earlier. In addition, Eska (2017) discusses some Celtiberian inscriptions that make a distinction between two stop series,[6] giving special attention to examples where sequences of a (historically voiceless) fricative and a stop are spelled with the LENIS stop grapheme ⟨sd⟩. Eska (2017) suggests that this reflects the lack of aspiration after a fricative, much as in present-day Welsh, Irish, and Scottish Gaelic (where similar spellings of such clusters are, or have been, attested in the written tradition).

A second source of evidence is the typology of sound changes affecting Proto-Celtic stops in the descendant languages. For instance, Eska (2018) considers the "loss" of Proto-Indo-European **p* in Celtic: as already noted, this process is normally considered to have proceeded via multiple stages involving voiceless fricatives (**p* > **ɸ* > **h* > ∅ *vel sim.*). Eska (2018) argues that this is compatible with the presence of aspiration on the stop, with reference to Kümmel (2007: 57). Further arguments of this kind are given in Eska (2018) with respect to the Common Celtic period and by Eska (2019; 2020) for the medieval and modern attestations. It is to a re-examination of this typological evidence that I now turn.

[6]The inscriptions in question are analysed by Jordán Cólera (2007); for a recent overview of Palaeohispanic writing systems, see Ferrer & Moncunill (2019).

4 Examining the sound changes

In this section, I examine the typology of sound changes affecting Proto-Celtic stops in the attested languages. After a brief overview of the overall pattern, I will consider the relevant changes by dividing them into four groups: changes that are more likely if Common Celtic belonged to the "voicing" type; changes more compatible with an "aspirating" system; changes where a reasonable, albeit non-trivial case can be made for either possibility; and finally non-probative changes that are easily compatible with either a "voicing" or an "aspirating" pattern in Common Celtic.

4.1 Lenition of stops in Gaelic and Brythonic

Processes of lenition affecting both FORTIS and LENIS stops are a hallmark of Celtic historical phonology. They have been the subject of extensive scholarship over the years, some of which is of foundational importance even outwith the specific concerns of Celtic studies (such as Martinet 1955). In this chapter, however, I am mostly concerned with what these changes reveal about the nature of laryngeal contrast prior to their operation.

The fundamental facts are fairly simple. In intervocalic (or, rather, postvocalic) position, both series of stops in the Insular Celtic languages are reflected with outcomes that are advanced along well-understood lenition trajectories relative to their original state. The developments are exemplified in Table 1. The consonants of interest are bolded, and the transcription of the relevant consonant given when not trivial. As it shows, LENIS stops changed to LENIS fricatives **β ð ɣ* in both Brythonic and Gaelic languages. In the case of FORTIS stops, there is a difference: in Gaelic, lenition produced FORTIS fricatives, whilst in Brythonic it resulted in LENIS stops. As for geminates, in the present-day languages LENIS geminates **bb dd gg* are (mostly) reflected as singleton LENIS stops, whilst FORTIS geminates again show a split: Brythonic reflects them as FORTIS fricatives **f θ x*, whereas in Gaelic they are reflected as FORTIS stops **p t k*.

Most of these processes are dated to some point in the sub-Roman period of the mid 1st millennium CE (Jackson 1953, Sims-Williams 1990, McCone 1996): they are generally not reflected (even incidentally) in the Continental material[7] or in the very earliest Romano-British elements borrowed into English. However, they do affect Latin borrowings in both Irish and Brythonic, as well as Brythonic and

[7] But see Section 4.5.1 on the LENIS stops and Stifter (2012) on the possibly related lenition of **s* to **h* in Gaulish.

Table 1: Postvocalic stops in Celtic

		Singletons		Geminates	
Language		FORTIS	LENIS	FORTIS	LENIS
Proto-Celtic		**katu-* 'battle'	**budaro-* 'deaf'	**(ɸ)likkā-* 'flat stone'	**kred-dī-* 'believe'
Gaelic	Old Irish	*ca**th*** [θ]	*bo**d**ar* [ð]	*le**cc*** [k]	*crei**t**id* [dʲ]
	Modern Irish	*ca**th***	*bo**dh**ar*	*lea**c***	*crei**d**im*
Brythonic	Welsh	*ca**d***	*by**dd**ar* [ð]	*lle**ch*** [χ]	*cre**d**u*
	Breton	*ka**d***	*bouzar*	*le**c'h***	*kre**d**iñ*

British Latin borrowings in Irish (McManus 1983), and were certainly complete by the time of the manuscript attestations.

4.2 Evidence for a "voicing" system

This category includes three unrelated changes: early "devoicing" of geminates, the Brythonic lenition of FORTIS stops to LENIS stops characteristic of Brythonic, and the development of FORTIS stops after nasals in Gaelic.

4.2.1 Geminate devoicing

The development of Common Celtic LENIS geminate stops **bb dd gg* presents difficult problems. They mostly arose from consonant clusters, usually at morpheme boundaries (Stifter 2024): Welsh (W) *credu* 'believe', Modern Irish (Ir) *creidim* from Proto-Celtic (PCelt) **kred-dī-* (cf. Sanskrit *śraddhấ*); W *aber* 'estuary' from **ad-bero-*, Old Irish (OIr) *acaldam* 'colloquy' (present-day *agallamh*) to the verb *ad:gládathar* 'address.prs.3sg'. In general, they are reflected in the present-day languages as short LENIS stops.

That said, there are least some examples of LENIS geminates, or LENIS consonant clusters, that have FORTIS outcomes in Brythonic (Pedersen 1909: 159–161; Jørgensen 2022, Stifter 2024: 145–146). Significantly, such examples are mostly root-internal: W *bychan*, Breton (Br) *bihan* 'small' (**bikko-*) compared to OIr *bec*, Ir *beag* (**biggo-*); Br *bouc'h* 'blunt' (**bukko-*) but OIr *boc* 'soft, tender', (Ir, ScG) *bog* (**buggo-*), and a few others. Also likely related is the reflex of the **sd* (> **zd*?) cluster as θ in cases like W *nyth* 'nest', (Middle Breton (MBret) *nez*, *nyth*, from

**nisdo-* (in contrast to OIr *net*, Ir *nead*). The simplest explanation of this development is devoicing of a voiced geminate in Brythonic: as Stifter (2024) points out, it is strongly supported by the **-tt-* appearing in Gallo-Romance borrowings reflecting Celtic **-sd-*, such as *pettia* 'part, piece' (French *pièce*) to (PCelt) **k^{w}esdi-* (*OIr cuit* 'share, part', Ir, ScG *cuid* but W *peth* 'thing', Br *pezh*). The same outcome is visible in Scottish placenames reflecting Pictish **pett-* as *Pet-*, *Pit-*.

This development is not entirely regular, and not all relevant consonants and clusters undergo it. In many morpheme-boundary cases, for instance after the prefix **ek(s)-*, such as W *eglan* 'seashore' (**ek-glannā-*), the context is highly vulnerable to cyclic/analogical effects (cf. unprefixed W *glan* 'shore'), and the data is notably difficult (Russell 1988, Stifter 2024). I am persuaded, however, by Jørgensen's (2022) argument that devoicing in items like **biggo-* was early, and the irregularities arise because it did not operate after the stabilization of LENIS geminates across historical morpheme boundaries, as in **ad-bero-* 'estuary' and **kred-dī-* 'believe'.

Typologically, devoicing of voiced geminate obstruents and clusters is strongly motivated phonetically (Ohala 1997) and well known typologically: for example, phonetic devoicing of such geminates occurs in Japanese (Hirose & Ashby 2007); and a diachronic/morphophonological pattern of this kind is found in Chaha (Banksira 2000: ch. 2). Under the aspiration-based account, the change from **bb dd gg* to **pp tt kk* must be reconstructed as *[tt] > *[t^{h}] (or even [tt^{h}]?), with subsequent spirantization. This, I suggest, is unlikely. Degemination of a plain stop to an aspirated one is not an impossible change: indeed, it is commonly cited as a possible source of preaspiration (Blevins & Garrett 1993). The rise of (post)aspirated geminates from plain ones is also attested, for example in Cypriot Greek (Newton 1972, Armostis 2009). However, the ultimate outcome of this change in the Brythonic languages is a fricative: neither preaspirated stops nor geminates of any kind are a plausible immediate source for a fricative.[8] I conclude that devoicing of voiced obstruent geminates/clusters is a far more preferable interpretation. (On the spirantization seen in Brythonic, see Section 4.4.2.)

4.2.2 Lenition of |fortis| stops in Brythonic

As we saw in Section 4.1, Gaelic and Brythonic differ in how lenition in (roughly) postvocalic position affects singleton FORTIS stops. Here, we focus on their reflexes in Brythonic, returning to the Gaelic pattern in Section 4.4.1. The outcomes

[8] In principle, one could imagine preaspirated stops undergoing preaspiration oralization (e. g. Silverman 2003) with subsequent loss of a stop (**tt* > **^{h}t* > **θt* > *θ*), but this very long sequence of events stretches credulity.

of original FORTIS stops in roughly postvocalic position are LENIS stops: W*cad* 'battle' < **katu-*.

This Brythonic "first lenition" is entirely unproblematic for a "voicing" language: intervocalic stop voicing is both phonetically and typologically trivial. It is more difficult to analyse it as involving loss of aspiration, as would be required under the alternative analysis. Eska (2020) proposes that this change resulted from earlier timing of the glottal spreading gesture in postvocalic position, leading to shorter VOT and reinterpretation of the relevant stops as LENIS. While not impossible, I suggest this explanation is problematic.

Voice onset time – the duration of postaspiration – is clearly an important cue to laryngeal contrast in word-initial position, and the change indeed applies to word-initial consonants in postvocalic sandhi (**esjō teɣos* 'his house' > Welsh *ei dŷ*). However, it is less clear that VOT is particularly important word-medially. To be sure, there are reports in the literature of languages with consistently postaspirated FORTIS stops word-medially: examples include German (Jessen 1998), Persian (Bijankhan & Nourbakhsh 2009), Northern Icelandic (Pétur Helgason 2002), and Georgian (Vicenik 2010). However, it is not clear what would motivate such a leftward shift in glottal opening after a vowel. Typologically, there are examples of across-the-board timing changes – consonant shifts – but it seems that the more usual case is an increase in VOT rather than a shortening. Grimm's Law in Germanic and the Armenian consonant shift (e. g. Sayeed & Vaux 2017) are celebrated instances in Indo-European, but we can find parallels elsewhere: for instance, a similar shift has occurred (apparently independently) in at least two branches of Austroasiatic: Angkuic and Mal-Pray (Sidwell 2021). However, it is more difficult to understand such a change as a positionally conditioned pattern with relatively earlier glottal opening specifically in leniting positions.

According to Kümmel (2007: 49), intervocalic lenition of FORTIS stops that is realised as loss of aspiration rather as voicing is something of a northern European specialty: his examples are primarily Germanic varieties such as Icelandic, Danish, and Western Norwegian. I cannot discuss these parallels in detail here, and refer the reader to Iosad (2025), where I suggest that this kind of lenition involves not a shortening of postaspiration, but reduction of preaspiration in FORTIS stops. It is conceivable, but difficult to prove, that a similar scenario applied in Celtic, too. This would give some support to its analysis as an "aspirating" language, although not in the exact way envisaged by Eska (2020).

There is a second possibility. A distribution whereby original FORTIS stops yield aspirated reflexes in "strong" positions, but unaspirated ones (ultimately merging with original LENIS ones) in "weak" positions could result not from

lenition targeting weak positions but from fortition in strong contexts. For example, Hill (2007) argues that Tibetan underwent aspiration of stops in initial position, which yields synchronic distribution analysable as intervocalic deaspiration in Old Tibetan (Hill 2010) or Amdo Tibetan (Green 2012). This possibility cannot be easily rejected for Brythonic. Indeed, it is also consistent with the fact that stops escape (first) lenition not only word-initially but also after a sonorant (W *gwy**nt*** 'wind', *se**rch*** 'love' with subsequent spirantization): it is widely recognized that the "coda mirror" position {#_, C_} often counts as "strong" (Ségéral & Scheer 2008). However, even if feasible, this scenario presupposes an *absence* of aspiration in FORTIS stops in strong position prior to this change, and thus is not compatible with a view of earlier Celtic as an aspirating language.[9]

To summarize, of the three possible scenarios that have been identified, two imply that Brythonic was a "voicing" language at least prior to the change, and of these one is phonetically trivial; only one is fully compatible with an aspirating system, but relies on rather tentative parallels. On balance, therefore, I suggest that the Brythonic first lenition is much more compatible with a "voicing" system of laryngeal contrast.

4.2.3 Nasal-stop clusters in Gaelic

A third development that strongly supports reconstructing an earlier "voiced" pattern is exclusive to Gaelic, where clusters of a (lenis) nasal followed by a FORTIS stop yield LENIS stops: **kantom* 'hundred', (OIr *cét*, Ir *céad*, ScG *ceud*, M *keead* (but W *cant* 'hundred'). This change has been interpreted in multiple ways, but I suggest that all of them are more compatible with a voicing contrast.

Traditional accounts (Thurneysen 1946, Martinet 1952; see also Stifter 2024) consider the development to have gone through a LENIS geminate stop: **kanto-* > **kēdd*. This interpretation disaggregates the change into two components, gemination and voicing. I will analyse them separately.

There are two main approaches to the voicing of the stop in the literature. One, endorsed by McCone (1996) under the rubric of "second lenition" (see also Kortlandt 1982), considers the voicing to result from postvocalic lenition that applied to newly postvocalic FORTIS stops after the loss of a preceding consonant: **kant-* > **kēt-* > **kēd-*. This is essentially the same process as the Brythonic "first lenition" (Section 4.2.2), and the same considerations apply to it: it is easy to account for under a "voicing" view, but more difficult, albeit not impossible, under an "aspiration" account.

[9] A similar scenario for the origin of the "aspirating" systems in the present-day Germanic languages is envisaged by Goblirsch (2005), building on Steblin-Kamenskij (1974).

A second view, also discussed by McCone (1996: 91) building on Greene (1960), is that the voicing was triggered by the nasal.[10] This proposition is much more difficult to reconcile with an original "aspirating" system. Nasals can, of course, exert a voicing influence on following stops, but this does not seem to commonly result in a neutralizing loss of aspiration. In Section 2.4, I discussed the postnasal voicing processes of Scottish Gaelic: generally, the outcomes of nasal + FORTIS stop sequences are distinct from underlying LENIS stops, and when they do result in a merger, it is precisely where the FORTIS stops show little if any aspiration.

A close parallel is provided by (Amur) Nivkh [gily1242]. It contrasts aspirated FORTIS stops [pʰ tʰ cʰ kʰ qʰ] with voiceless unaspirated LENIS stops [p t c k q], and possesses a postnasal voicing process quite similar to the Scottish Gaelic "eclipsis". Crucially for our purposes, it affects LENIS (unaspirated) stops, causing them to become voiced (*təf-**k**u* 'house-PL' but *qan-**g**u* 'dog-PL', but fails to affect FORTIS (aspirated) stops (*alʂ-**p**ʰe* 'pick berries' but *tʰoʁzaŋ-**p**ʰe* 'pick nuts').[11] More generally, Kümmel (2007: 54) finds that postnasal voicing of aspirated stops usually produces aspirated or breathy-voiced outcomes. If anything, postnasal position can be associated with plain stops becoming aspirated, as attested in numerous Bantu languages (Downing & Hamann 2018).[12] I conclude that a **nt* > **(n)d* scenario is much more strongly compatible with a view where the contrast is of the "voicing" type.

The second potential component of the change, gemination, is perhaps less relevant for our concerns (see Section 4.5.2 below). The literature widely accepts it to have occurred in these clusters, even though the evidence is rather meagre. If gemination did occur, it could be reconstructed as assimilation of **nd* clusters produced by postnasal voicing (this finds a good parallel, for instance, in a number of Semitic languages), or a direct **nt* > **dd* change, with a parallel in the Western Sámi languages (Sammallahti 1998; see further Kümmel 2007: 180–181). As far as can be seen, the outcomes of this process merge with original degeminated LENIS stops, as in OIr *bec* 'small' (**biggo-*), but the dating of the degemination is unclear: as long as it follows the (likely very early) spirantization of original **b d g* (see Section 4.5.1 below), it could be dated to almost any time in

[10] Such clusters do not merge with original **nd*-type sequences, because in the latter the nasal became FORTIS and was preserved: **angu-* > **ɛNgu-* > *ing* 'straight', in contrast to **anku-* > **ɛngu-* > *éc* 'death'.

[11] Examples from Gruzdeva (2024: 502).

[12] Hyman (2001) notes a synchronic postnasal deaspiration process in the Nguni languages [ngun1276], but diachronically it represents the blocking of aspiration of plain voiceless stops after nasals. If anything, this pattern *supports* the proposition that FORTIS stops were originally plain.

the history of the Gaelic languages. Whether or not the original nasal-stop clusters went through a geminate stage does not influence my conclusions as to the nature of laryngeal contrast.

4.3 Evidence for an "aspirating" system

Two sound changes provide strong support for an "aspirating" system of laryngeal contrast. They are both late in the relative chronology, because they specifically affect "secondary" clusters that arose via a process of unstressed vowel syncope. This syncope can be dated to the early to mid 6th century CE (Jackson 1953, Sims-Williams 1990).

4.3.1 Voicelessness assimilation

In clusters created by syncope where the second member was a FORTIS obstruent (not least [h]), one can observe regressive laryngeal assimilation. These processes occur in both Brythonic and Gaelic. For Brythonic examples, see W *ateb* 'answer' from **ad-heb-* < **ati-sek^{w}-*, *drycin* 'storm' < **drug-hin-* < **druko-sīn-*. In Gaelic, there are examples like OIr *de**phth**igim* 'fight.PRS.1SG', Ir *dea**f**ach* 'contentious' (OIr *de**bth**ach*) by contrast with *de**b**uith* 'discord', Ir *dea**bh**aidh* (McCone 1996: 128). For the behaviour of *h* in Gaelic, cf. *im**b:s**oí* 'turn.PRS.3SG' but prototonic *-im**p**ai* from **ambi-sowet(i)*. As I discussed in Section 2.1, this kind of pattern, which privileges FORTIS phonological activity, is very strongly compatible with (if not diagnostic for) an aspirating system.

4.3.2 Provection

A second post-syncope development, relevant only for Brythonic, also concerns secondary obstruent clusters: LENIS geminates arising from syncope yield FORTIS stops in the present-day languages, as in Welsh *llety* 'lodging' from **led-deɣ-*.[13] Under an account like Eska's (2020), in which the LENIS stops were voiceless, the development, commonly known as "provection", can be interpreted as a change from a voiceless geminate to a (preaspirated) stop, which I already noted to be typologically well supported. Under an account based on voicing, the development needs to proceed via an intermediate stage of geminate devoicing: this is, as already noted, unproblematic phonetically, but makes the derivation more complex. Overall, this process does not appear to be fully probative, but an aspirating analysis is more parsimonious in this case.

[13] Cf. also the placenames Lichfield (Staffordshire) and Lytchett (Dorset) from Brythonic **lēd-gēd-* < **lēto-kaito-* 'grey forest' (Jackson 1953: 563).

4.4 Ambiguous changes: Spirantization of FORTIS stops

One change that has a rather ambiguous status is the spirantization of FORTIS **(p) t k* stops to the FORTIS non-strident fricatives **f θ x*. It occurs separately, with different conditioning, in the two branches.

4.4.1 Goidelic first lenition

As discussed in Section 4.1, FORTIS fricatives **θ x* x^w are the usual outcome of original singleton FORTIS stops in Goidelic in postvocalic position (including in postvocalic sandhi). The usual explanation is that the stops were originally aspirated and became fricatives likely via an affricate stage, a change with multiple typological parallels (not least the High German Consonant Shift); see Koch (1990), Isaac (2004). Eska (2018) further assimilates the spirantization of Proto-Indo-European **p* to **ɸ* in Early Celtic to this parallel, stretching the possible presence of aspiration even earlier.

In support of this proposition, Eska (2018) cites the findings of Kümmel (2007: 57). However, these results warrant further scrutiny: what Kümmel (2007) finds is that aspiration is necessary for FORTIS spirantization when it is an across-the-board consonant shift, but both changes of individual consonants (as with PCelt *ɸ* < PIE **p*) and positionally restricted lenition processes are attested as spirantizations of what are unaspirated FORTIS stops. One particularly well-known example of the latter is *gorgia toscana*, the postvocalic lenition of singleton FORTIS stops found in Tuscan Italian [ital1282]. In a recent in-depth discussion of this process, Marotta (2008) argues that it does not proceed via affrication of aspirated stops, although see Honeybone (2023) for some skepticism on this possibility. Another possible case is reported from Southern Jutland Danish by Puggaard-Rode (2024).

My conclusion is that the spirantization of FORTIS stops cannot clearly distinguish between the two approaches to laryngeal contrast. It is more easily compatible with an "aspirating" system, but a "voicing" pattern cannot be conclusively ruled out at this stage.

4.4.2 Brythonic second lenition

In Brythonic, FORTIS stops have fricative outcomes in two contexts. First, this is observed regularly with postvocalic FORTIS geminates: W *llech* 'slate', Br *lec'h*, Cornish *lehan* < **(ɸ)likkā*. Second, the same outcome is found for FORTIS singletons in two contexts: after a liquid (*serch* < **sterkā* 'love'; contrast OIr *sercc*, Ir *searc*) and in sandhi after the loss of certain word-final consonants, especially **h*

from earlier **s* (*esjas teɣos* 'her house' > **ehja θeɣoh* > W *ei thŷ*; contrast OIr *a (t)tech*, Ir *a teach*).

A long line of work (Jackson 1953, 1960, Koch 1989, 1990, Isaac 2004, 2008) has taken this development to support the presence of aspiration on FORTIS stops: the greater "force" of geminate stops is hypothesized to have facilitated aspiration, followed by affrication and subsequent spirantization (**tt^h* > **t(t)θ* > *θ*).

However, authors such as Thomas (1990), Sims-Williams (1990, 2008) and McCone (1996) have suggested this is unnecessary: instead, they analyse the Brythonic developments as a combination of FORTIS degemination (non-probative for our purposes, Section 4.5.2) followed by spirantization of the type discussed in Section 4.4.1 (so **likkā-* > **likā-* > *llech*). I find the latter scenario more persuasive. In particular, the direct development from geminates to fricatives without degemination is typologically highly suspect (e. g. Kirchner 1998). This analysis also allows us to dispense with the otherwise highly odd gemination that the traditional scenario has to posit to account for spirantization of singletons in forms like W *serch*, *ei thŷ*. Instead, all of these developments can be subsumed under a "second lenition", where postvocalic singleton FORTIS stops of secondary origin (i. e. produced by degemination or by loss of a preceding consonant), as well as post-liquid stops, are reflected as FORTIS fricatives, parallel to the "first lenition" in Gaelic. If one accepts the degemination-and-lenition scenario, the Brythonic development is now a combination of two non-probative changes.

4.5 Non-probative changes

Finally, I briefly consider two categories of changes that are not probative for the nature of the Early Celtic system, being easily compatible with either option.

4.5.1 Spirantization of |lenis| stops

Original LENIS singleton stops yield LENIS fricatives, generally reconstructed as **β ð ɣ*, in both Gaelic and Brythonic. They are mostly reflected as fricatives, usually (but sometimes variably) voiced, in the present day (Ball & Williams 2001). Intervocalically, they often undergo further lenition processes such as debuccalization and ultimately deletion, especially in Irish and Breton, but the general pathway is widely agreed on.

Although traditionally this process is dated to the same mid 1st millennium CE stage as the other Insular lenitions, more recent scholarship (e.g. Sims-Williams 1990; Villar 1993; McCone 1996; Eska 2017) has been willing to date it to a much

earlier period. The reason is that there are some traces of this process (such as loss of intervocalic **g* and fricative spellings of postvocalic **d*) in both Gaulish and Celtiberian.

This change is not especially informative. Voiced fricatives are commonly produced by lenition of voiced stops, but they can also arise from the spirantization of voiceless, contrastively unaspirated stops. *Comparanda* for the lenition of voiced stops to voiced fricatives are trivial to find: for instance, this process is pervasive in Romance languages, including Spanish [stan1288], Catalan, and many languages of Italy. However, clearly "aspirating" languages can also show a pattern of LENIS voiceless unaspirated stops alternating with voiced continuants. I have already cited Nivkh. The two series of stops, both aspirated /pʰ tʰ cʰ kʰ qʰ/ and voiceless unaspirated /p t c k q/, participate in a morphologically restricted process of morpheme-initial consonant mutation. They alternate with voiceless and voiced continuants [f ɻ̊ s x χ] and [v r z ɣ ʁ] respectively, just like in Scottish Gaelic: *təf-**k**u* 'house-PL' but *mu-**ɣ**u* 'boat-PL'. I conclude that the lenition of singleton LENIS stops to voiced fricatives, whether treated as part of the Insular lenition processes or put further back in time, does not furnish probative arguments regarding the nature of laryngeal contrast at the time that it occurred.

4.5.2 Degemination

The clearest case of degemination is found with FORTIS geminates in the Gaelic languages, where they are reflected as FORTIS stops (with, where appropriate, preaspiration): **(ɸ)likkā-* 'flat stone' > IR, ScG, *leac*, M *lhiack*. There is no contrastive consonant (or in any case stop) length in the present-day languages, so these stops are generally considered to be short (with a few exceptions). This implies a degemination at some point along the way. Unfortunately, there is no consensus as to when precisely it occurred: the possibilities range from the date prior to the manuscript record to, in some cases, incomplete degemination up to the present day (on issues of consonant quantity in Gaelic, see recently Lewin 2023, Stifter 2024, Wheatley & Iosad in preparation). Whatever the case may be, degemination itself is a trivial change that does not offer a criterion to distinguish between the possible systems of laryngeal contrast.

5 Reconsidering the evidence

Having evaluated relevant changes individually, I can now synthesize the evidence for the two branches. In both cases, I will argue that the earlier changes

are overall more compatible with a voicing-based system, while syncope marks the advent of the present-day "aspirating" patterns.

5.1 Gaelic

Table 2 summarizes the developments in the Gaelic languages. To facilitate interpretation, cells corresponding to strongly diagnostic changes are shaded. The degemination of FORTIS stops is difficult to place in the chronology, and it is in any case not probative, so it is excluded. As for LENIS geminates, one could perhaps place the degemination of the original ones relatively early, but since any geminates that arose due to syncope appear to have been shortened on par with the earlier ones, for the sake of concreteness I show a single post-syncope round of degemination (in agreement with Stifter 2017: 1199).

Table 2: Consonant developments in Gaelic

Early changes	Voicing scenario	Aspiration scenario
(p) t k > (f) θ x / _V **katu- > cath*	Compatible	Strongly compatible
(p) t k > (b) d g / _N **kant- > cét*	Strongly compatible	Weakly or not compatible
Post-syncope changes		
bb dd gg > b d g **biggo- > bec* **ag-gald- > acaldam*	Not probative	Not probative
Assimilation **-ámbi-sowet(i) > -impai*	Unlikely	Strongly compatible

My interpretation of Table 2 is that the earlier sound changes are very strongly compatible with a "voicing" system of laryngeal contrast, whereas an "aspirating" system is only very weakly, if it all, compatible with the data. The behaviour of nasal-stop clusters in particular is very difficult to square with an "aspirating" analysis. Admittedly, the evidence for spirantization of unaspirated voiceless stops is problematic in some ways, but the phenomenon appears to be not entirely unattested.

On the other hand, the later, post-syncope pattern is essentially what is found in the present-day languages, which show prototypical "aspirating" behaviour.

Therefore, the prudent assumption is that secure evidence for the "aspirating" system does not date to earlier than around the time of syncope, late in the Primitive Irish period, rather than to an earlier period in the history of Celtic, *pace* Eska (2018).

I suggest there is some circumstantial evidence that supports reconstructing the rise of aspiration in Gaelic FORTIS stops to the late Primitive Irish period. Almost all present-day Gaelic languages show a process of vowel intrusion in heterorganic sequences of a sonorant and another consonant (as in Irish [dʲarəg] *dearg* 'red', OIr *derc*); see Morrison (2019), Iosad & Maguire (in preparation) for details. Today, this process is regularly blocked when the second consonant in such a cluster is a FORTIS stop: such clusters lack intrusive vowels and in many dialects, especially in Scottsh Gaelic, voiceless sonorants are found instead ([ɔl̥k] *olc* 'evil'). Morrison (2019) persuasively argues that the intrusive vowel arises whenever the sonorant is moraic in the stem-level phonology. If this is correct, the FORTIS stop exception can be understood as failure of the sonorant to acquire a mora before such stops. Morrison (2019) points to East Perthshire Gaelic (Ó Murchú 1989), where the synchronic generalization is that coda sonorants are long (thus moraic) before historically LENIS, but not historically FORTIS stops ([palːk] *balg* 'bag' but [ɔlk] *olc* 'evil'), a pattern Morrison (2019) considers archaic. One can hypothesize that sonorants before FORTIS stops fail to acquire a mora precisely because they were devoiced by the following stop (on the connection between voicing and moraicity, see Zec 1988), a phenomenon closely allied to preaspiration.

Now, although the exceptions to the generalization that FORTIS stops block vowel intrusion are marginal in the present-day languages, Eska (2010) identified a number of Ogam inscriptions that witness what he calls "non-canonical" vowel intrusion, with examples such as ⟨ERACIḌAS⟩ (cf. ⟨ERCIAS⟩ and later *Ercc*). He argues, convincingly, that, *pace* O'Rahilly (1936), Hind (1996), such spellings do reflect the presence of intrusive vowels in these contexts in Primitive Irish. If he is correct, then one can posit that intrusion was able to apply before FORTIS stops precisely because such stops did not trigger sonorant devoicing. Once devoicing set in, it blocked sonorant moraicity, and the intrusive vowels were suppressed. Given that we observe the intrusive vowels in this context in the Ogam corpus, but not in the later Gaelic languages, I suggest that the rise of sonorant devoicing, associated with aspiration, may be tentatively dated to the late Primitive Irish period – exactly as suggested by the analysis of Table 2.

5.2 Brythonic

Turning now to Brythonic, Table 3 shows the sequence of changes and their evaluations. I place the degemination of stops early in the chronology, because in the case of FORTIS stops it feeds second lenition, whilst LENIS stops degeminated prior to syncope, since they escape the devoicing experienced by secondary LENIS clusters.

Table 3: Consonant developments in Brythonic

Early changes	Voicing scenario	Aspiration scenario
p t k > b d g / _V **katu-* > *cad*	Strongly compatible	Weakly compatible
Sporadic geminate devoicing **bigg-* > *bikk-*	Strongly compatible	Highly unlikely
Degemination **bikk-an-* > **bikan-* > *bychan* **likkā-* > **likā-* > *llech* **ad-bero-* > **abbero-* > *aber*	Not probative	Not probative
p t k > f θ x / _V **broko-* > *broch* **sterkā* > *serch*	Compatible	Strongly compatible
Post-syncope changes		
bb dd gg > p t k **leto-teɣos* > **led-deɣ-* > *llety*	Strongly compatible but more complex	Strongly compatible
Assimilation **ati-sek^{w}-* > **ad-heb* > *ateb*	Unlikely	Strongly compatible

As in Gaelic, later developments, like the present-day patterns, agree strongly with the "aspirating" analysis. For the earlier stages, it would appear that both analyses are feasible. The balance might be tilted in favour of the voicing approach (as for Gaelic) if one considers that lenition as deaspiration (or preaspiration shortening) is somewhat rare typologically, and accepts the probative value of sporadic degemination despite its lack of regularity.

It is telling that the very earliest changes in Brythonic are also the ones most strongly pointing towards a "voicing" analysis. Just as with Gaelic, the closer

one gets to the present day, the more the balance tips towards "aspirating" patterns, even though (with the problematic exception of geminate devoicing) a very strongly diagnostic pattern like the postnasal voicing of Gaelic cannot be found.

Overall, I conclude that while it is *possible* that FORTIS stops were aspirated at a very early stage of Celtic, as argued by Eska (2017; 2018; 2019; 2020), the evidence from the earliest sound changes is also compatible with a view of FORTIS stops as voiceless (likely unaspirated) and LENIS stops as voiced. The change to the present-day system can be tentatively dated to some point late in the era of considerable upheaval in the structure of the Insular Celtic languages in the mid 1st millennium CE.

6 Conclusions

Overall, I have argued that while most modern languages (other than Breton) are classic "aspirating" systems in the sense posited by Laryngeal Realism (in agreement with Eska 2018), solid evidence for these patterns does not stretch much further back than the sub-Roman period in Britain and Ireland, when the Celtic languages underwent a series of drastic phonological and morphological changes that have largely defined their shape in the medieval and modern period. In this final section, I offer some possible perspectives on the wider relevance of this conclusion for Celtic historical linguistics.

In particular, I suggest that the proposed interpretation does not necessarily mean that Early Celtic was necessarily a "voicing" language *tout court*. Rather, I suggest that we should take seriously the proposition that the Celtic languages, agreed to have been spoken over a large part of Western Europe throughout the 1st millennium BCE, demonstrated a degree of variability in their laryngeal contrast. Although the general slant of the reconstruction proposed here is that the development went from a "voicing" towards an "aspirating" system, the change need not have happened only once and/or in the same place.

Instead, I suggest that it is wise to envisage a pool of synchronic variation (Ohala 1989) in the realisation of laryngeal contrasts across prehistorical Celtic varieties, with different dialects undergoing often similar shifts at different points in time. Such a régime would explain, for example, the evidence discussed in Section 3, which seems to suggest an "aspirating" system being attested on the Continent several centuries before my proposed change in the Insular varieties. If "aspirating" patterns did indeed stabilize earlier in Continental Celtic, it may not be a coincidence that it was precisely the Continental varieties that were located near the spatial interface with Germanic (for up-to-date discussion of

Celtic-Germanic relationships, see van Sluis et al. 2023), widely accepted to have been just such an "aspirating" language from a very early stage (Salmons 2020).

Even within Insular Celtic, it is not clear that the change happened at the same time. Even if a very late date for the split of Goidelic from Brythonic is accepted (perhaps along the lines of Schrijver 2015), it would have already occurred by the time of syncope, a likely *terminus post quem* at least for the Gaelic languages. The shift towards an "aspirating" system may thus have been at least partly independent in the two Insular groups. In this way, this change presents another example of Sapirean drift driven by the narrowing down of the range of inherited variation (Joseph 2006, Iosad 2025).

A second corollary is that if some version of a "voicing" pattern was maintained at least in some part of the Celtic *Sprachraum* as late as the first millennium CE, then its occurrence in Breton does not necessarily have to be fully ascribed to the influence of Romance. If some aspects of a "voicing" system were maintained during the spread from Britain, then perhaps the role of contact could have been more complex, which may have consequences for our understanding of the sociolinguistics of Romance-Celtic interaction in Brittany.

More widely, I suggest that a model that takes synchronic variation in proto-languages seriously (cf. Natvig & Salmons 2020) is potentially productive in tackling many classic issues in Celtic historical linguistics. I leave additional elaboration of such a model to further research.

Abbreviations

Br	Breton	PCelt	Proto-Celtic
Ir	Modern Irish	PIE	Proto-Indo-European
M	Manx	ScG	Scottish Gaelic
MBret	Middle Breton	SGDS	Survey of the Gaelic Dialects of Scotland
MCo	Middle Cornish		
OIr	Old Irish	W	Welsh

Acknowledgments

I have been thinking about the subject of this paper for a long time, all the way back to 2006 when I first presented some fledgling thoughts at the Second Colloquium of Societas Celto-Slavica (Institute of Linguistics, Russian Academy of Sciences). I am grateful to the organizers of the Foundational Approaches to Celtic Linguistics series at the University of Arizona for the opportunity to present this

work in March 2023, and to the audience for stimulating questions. Thanks also to members of the Historical Phonology Reading Group at the University of Edinburgh for comments on a draft of this paper, and to two anonymous reviewers for numerous suggested improvements. All mistakes and shortcomings remain my own.

References

Allen, Liradie. 2022. *Plosive class distinctions in unpreaspirating varieties of Scottish Gaelic: A contribution to the study of SGDS*. Edinburgh: The University of Edinburgh. (MA thesis).

Armostis, Spyros. 2009. *The phonetics of plosive and affricate gemination in Cypriot Greek*. Cambridge: University of Cambridge. (Doctoral dissertation).

Ball, Martin J. & Briony Williams. 2001. *Welsh phonetics* (Welsh Studies 17). Lewiston: Edwin Mellen Press.

Banksira, Degif Petros. 2000. *Sound mutations: The morphophonology of Chaha*. Amsterdam: John Benjamins.

Beckman, Jill, Michael Jessen & Catherine Ringen. 2013. Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics* 49(2). 259–284. DOI: 10.1017/S0022226712000424.

Bell, Elise, Diana B. Archangeli, Skye J. Anderson, Michael Hammond, Peredur Webb-Davies & Heddwen Brooks. 2023. Northern Welsh. *Journal of the International Phonetic Association* 53(2). 487–510. DOI: 10.1017/s0025100321000165.

Bijankhan, Mahmood & Mandana Nourbakhsh. 2009. Voice onset time in Persian initial and intervocalic stop production. *Journal of the International Phonetic Association* 39(3). 335–364. DOI: 10.1017/s0025100309990168.

Blevins, Juliette & Andrew Garrett. 1993. The evolution of Ponapeic nasal substitution. *Oceanic Linguistics* 32(3). 199–238.

Bosch, Anna & James Scobbie. 2009. Fine-grained morpho-phonological variation in Scottish Gaelic: Evidence from the Linguistic Survey of Scotland. In James N. Stanford & Dennis R. Preston (eds.), *Variation in indigenous minority languages* (IMPACT: Studies in Language and Society 25), 347–368. Amsterdam: John Benjamins.

Cyran, Eugeniusz. 2013. *Polish voicing: Between phonology and phonetics*. Lublin: Wydawnictwo KUL.

Downing, Laura J. & Silke Hamann. 2018. The phonetics of NCh in Tumbuka and its implications for diachronic change. *Papers in Historical Phonology* 3. 77–95. DOI: 10.2218/pihph.3.2018.2825.

Eska, Joseph F. 2010. Remarks on intrusive vowels in the Ogam corpus of Early Irish. *Keltische Forschungen* 5. 139–156.
Eska, Joseph F. 2017. The significance of śDAḿ and related matters. *International Journal of Diachronic Linguistics and Linguistic Reconstruction* 14(1). 51–76.
Eska, Joseph F. 2018. Laryngeal realism and the prehistory of Celtic. *Transactions of the Philological Society* 116(3). 320–331. DOI: 10.1111/1467-968x.12122.
Eska, Joseph F. 2019. Laryngeal realism and early Insular Celtic orthography. *North American Journal of Celtic Studies* 3(1). 1–17. DOI: 10 . 26818 / nortamerceltstud.3.1.0001.
Eska, Joseph F. 2020. Interarticulatory timing and Celtic mutations. *Journal of Celtic Linguistics* 21. 235–255. DOI: 10.16922/jcl.21.7.
Ferrer, Joan & Noemí Moncunill. 2019. Palaeohispanic writing systems: Classification, origin, and development. In Alejandro G. Sinner & Javier Velaza (eds.), *Palaeohispanic languages and epigraphies*, 78–108. Oxford: Oxford University Press.
Goblirsch, Kurt Gustav. 2005. *Lautverschiebungen in den germanischen Sprachen*. Heidelberg: Universitätsverlag Winter.
Gray, Louis H. 1944. Mutation in Gaulish. *Language* 20(4). 223–230.
Green, R. Jeffrey. 2012. The phonology of voicing and aspiration in Amdo Tibetan. *Linguistics of the Tibeto-Burman Area* 35(2). 1–31. DOI: 10.32655/ltba.35.2.01.
Greene, David. 1960. Some problems of Irish phonology. *Celtica* 5. 103–106.
Gruzdeva, Ekaterina. 2024. The Amuric language family. In Edward Vajda (ed.), *The languages and linguistics of Northern Asia*. Vol. 1: *Language families* (World of Linguistics 10), 481–540. Berlin: Mouton de Gruyter. DOI: 10 .1515/ 9783110556216-009.
Hall, Tracy Alan. 2001. Introduction: Phonological representations and phonetic implementation of distinctive features. In Tracy Alan Hall (ed.), *Distinctive feature theory* (Phonetics and Phonology 2), 1–40. Berlin: Mouton de Gruyter.
Henderson, George. 1903. The Gaelic dialects (continuation). *Zeitschrift für celtische Philologie* 4. 493–524.
Hill, Nathan W. 2007. Aspirated and unaspirated voiceless consonants in Old Tibetan. *Language and Linguistics* 8(2). 471–493.
Hill, Nathan W. 2010. An overview of Old Tibetan synchronic phonology. *Transactions of the Philological Society* 108(2). 110–125. DOI: 10.1111/j.1467-968x.2010. 01234.x.
Hind, Kevin. 1996. The structure of epenthesis in Gaelic. *Journal of Celtic Linguistics* 5. 91–119.

Hirose, Aki & Michael Ashby. 2007. An acoustic study of devoicing of the geminate obstruents in Japanese. In *Proceedings of ICPhS XVI*, 909–912. Saarbrücken.

Honeybone, Patrick. 2005. Diachronic evidence in segmental phonology: The case of obstruent laryngeal specification. In Marc van Oostendorp & Jeroen van de Weijer (eds.), *The internal organization of phonological segments* (Studies in Generative Grammar 77), 319–354. Berlin: Mouton de Gruyter.

Honeybone, Patrick. 2008. Lenition, weakening and consonantal strength: Tracing concepts through the history of phonology. In Joaquim Brandão de Carvalho, Tobias Scheer & Philippe Ségéral (eds.), *Lenition and fortition* (Studies in Generative Grammar 99), 9–93. Berlin: Mouton de Gruyter.

Honeybone, Patrick. 2023. Can fortis stops spirantise without aspiration? Presentation at the 26th International Conference on Historical Linguistics, University of Heidelberg.

Hyman, Larry M. 2001. The limits of phonetic determinism in phonology: *NC revisited. In Elizabeth V. Hume & Keith Johnson (eds.), *The role of speech perception in phonology*, 141–186. San Diego: Academic Press.

Iosad, Pavel. 2017. *A substance-free framework for phonology: An analysis of the Breton dialect of Bothoa* (Edinburgh Studies in Theoretical Linguistics 2). Edinburgh: Edinburgh University Press.

Iosad, Pavel. 2023. The [ATR]/Laryngeal connection and emergent features. In Bert Botma, Florian Breit, Marc van Oostendorp & Marijn van 't Veer (eds.), *Primitives of phonological structure* (Oxford Studies in Phonology and Phonetics), 161–208. Oxford: Oxford University Press.

Iosad, Pavel. 2025. *Phonological drift and language contact: The northwest European phonological area: The northern European phonological area.* Cambridge: Cambridge University Press.

Iosad, Pavel. Forthcoming. Modern Breton phonology. In Joseph F. Eska, Silva Nurmio, Peadar Ó Muircheartaigh & Paul Russell (eds.), *The Palgrave handbook of Celtic linguistics.*

Iosad, Pavel & Warren Maguire. In preparation. English epenthesis in *lC* and *rC* clusters: Areal effect or drift? MS., University of Edinburgh.

Isaac, Graham R. 2004. The chronology of the development of Brittonic stops and the spirant mutation. *Journal of Celtic Linguistics* 8. 49–85.

Isaac, Graham R. 2008. Brythonic voiceless spirants again. *Journal of Celtic Linguistics* 12. 17–37.

Jackson, Kenneth Hurlstone. 1953. *Language and history in early Britain.* Edinburgh: Edinburgh University Press.

Jackson, Kenneth Hurlstone. 1960. Gemination and the spirant mutation. *Celtica* 5. 127–134.

Jessen, Michael. 1998. *Phonetics and phonology of tense and lax obstruents in German*. Amsterdam: John Benjamins.

Jones, Glyn E. 1984. The distinctive vowels and consonants of Welsh. In Martin J. Ball & Glyn E. Jones (eds.), *Welsh phonology: Selected readings*, 40–64. Cardiff: University of Wales Press.

Jordán Cólera, Carlos. 2007. Estudios sobre el sistema dual de escritura en epigrafía no monetal celtibérica. *Palaeohispanica* 7. 101–142. DOI: 10 . 36707 / palaeohispanica.v0i7.278.

Jørgensen, Anders Richardt. 2022. Celtic. In Thomas Olander (ed.), *The Indo-European language family: A phylogenetic perspective*, 131–151. Cambridge: Cambridge University Press.

Joseph, Brian D. 2006. On projecting variation back into a proto-language, with particular attention to Germanic evidence and some thoughts on "drift". In Thomas D. Cravens (ed.), *Variation and reconstruction* (Current Issues in Linguistic Theory 268), 103–118. Amsterdam: John Benjamins. DOI: 10.1075/cilt.268.06jos.

Kirby, James & D. Robert Ladd. 2019. Effects of obstruent voicing on vowel F_0: Implications for laryngeal realism. *Yearbook of the Poznań Linguistic Meeting* 4. 213–235. DOI: 10.2478/yplm-2018-0009.

Kirchner, Robert. 1998. *An effort-based approach to consonant lenition*. Los Angeles: University of California, Los Angeles. (Doctoral dissertation).

Koch, John T. 1989. Neo-Brittonic voiceless spirants from Old Celtic geminates. *Ériu* 40. 119–128.

Koch, John T. 1990. **Cothairche*, Esposito's theory, and neo-Celtic lenition. In Alfred Bammesberger & Alfred Wollmann (eds.), *Britain 400–600: Language and history*, 97–119. Heidelberg: Carl Winter Universitätsverlag.

Kortlandt, Frederik. 1982. Phonemicization and rephonemicization of the Old Irish mutations. *Ériu* 33. 73–83.

Kümmel, Martin. 2007. *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen für die vergleichende Rekonstruktion*. Wiesbaden: Dr. Ludwig Reichert Verlag.

Ladefoged, Peter, Jenny Ladefoged, Alice Turk, Kevin Hind & St. John Skilton. 1998. Phonetic structures of Scottish Gaelic. *Journal of the International Phonetic Association* 28(1). 1–41. DOI: 10.1017/S0025100300006228.

Lewin, Christopher. 2023. Preocclusion in Manx. *Journal of Celtic Linguistics* 24(1). 125–166. DOI: 10.16922/jcl.24.5.

Lewis, Henry & Holger Pedersen. 1937. *A concise comparative Celtic grammar*. Göttingen: Vandenhoeck & Ruprecht.

Marotta, Giovanna. 2008. Lenition in Tuscan Italian (gorgia toscana). In Joaquim Brandão de Carvalho, Tobias Scheer & Philippe Ségéral (eds.), *Lenition and fortition* (Studies in Generative Grammar 99), 235–271. Berlin: Mouton de Gruyter.

Martinet, André. 1952. Celtic lenition and Western Romance consonants. *Language* 28(2). 192–217.

Martinet, André. 1955. *Économie des changements phonétiques*. Bern: Francke.

McCone, Kim. 1996. *Towards a relative chronology of Ancient and Medieval Celtic sound change*. Kindare: Maynooth.

McManus, Damian. 1983. A chronology of the Latin loan-words in Early Irish. *Ériu* 34. 21–71.

Morris, Jonathan & Míša Hejná. 2020. Pre-aspiration in Bethesda Welsh: A sociophonetic analysis. *Journal of the International Phonetic Association* 50(2). 168–192. DOI: 10.1017/S0025100318000221.

Morrison, Donald A. 2019. Metrical structure in Scottish Gaelic: Tonal accent, glottalisation and overlength. *Phonology* 36(3). 391–432. DOI: 10.1017/S0952675719000204.

Nance, Claire & Roibeard Ó Maolalaigh. 2021. Scottish Gaelic. *Journal of the International Phonetic Association* 51(2). 261–275. DOI: 10.1017/s002510031900015x.

Nance, Claire & Jane Stuart-Smith. 2013. Pre-aspiration and post-aspiration in Scottish Gaelic stop consonants. *Journal of the International Phonetic Association* 43(2). 129–152. DOI: 10.1017/S0025100313000042.

Natvig, David & Joseph C. Salmons. 2020. Fully accepting variation in (pre)history: The pervasive heterogeneity of Germanic rhotics. In Patricia C. Sutcliffe (ed.), *The polymath intellectual: A Festschrift in honor of Robert D. King*, 81–102. Dripping Springs: Agarita Press.

Newton, Brian. 1972. *Cypriot Greek: Its phonology and inflections* (Janua Linguarum Series Practica 121). The Hague: Mouton.

Ní Chasaide, Ailbhe. 1985. *Preaspiration in phonological stop contrasts: An instrumental phonetic study*. Bangor: University College of North Wales. (Doctoral dissertation).

Ó Buaachalla, Breandán. 1985. The f-future in Modern Irish: A reassessment. *Proceedings of the Royal Irish Academy, Section C* 85C. 1–36.

Ó Dochartaigh, Cathair (ed.). 1994. *Survey of the Gaelic dialects of Scotland*. Dublin: Dublin Institute for Advanced Studies.

Ó Maolalaigh, Roibeard. 1995. The development of eclipsis in Gaelic. *Scottish Language* 14–15. 158–173.

Ó Murchú, Máirtín. 1985. Varia VIII: Devoicing and pre-aspiration in varieties of Scots Gaelic. *Ériu* 36. 195–198.
Ó Murchú, Máirtín. 1989. *East Perthshire Gaelic: Social history, phonology, texts and lexicon*. Dublin: Dublin Institute for Advanced Studies.
Ó Raghallaigh, Brian. 2014. *Fuaimeanna na Gaeilge*. 2nd edn. Baile Átha Cliath: Cois Life.
O'Rahilly, Thomas F. 1936. *Irish dialects: Past and present, with chapters on Scottish and Manx*. Dublin: Browne & Nolan.
Ohala, John J. 1989. Sound change is drawn from a pool of synchronic variation. In Leiv Egil Breivik & Ernst Håkon Jahr (eds.), *Language change: Contributions to the study of its causes* (Trends in Linguistics: Studies and Monographs 43), 173–198. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110853063.173.
Ohala, John J. 1997. Aerodynamics of phonology. In *Proceedings of the 4th Seoul International Conference on Linguistics*, 92–97. Seoul: Linguistic Society of Korea.
Pedersen, Holger. 1909. *Vergleichende Grammatik der keltischen Sprachen*, vol. 1. Göttingen: Vandenhoeck & Ruprecht.
Pétur Helgason. 2002. *Preaspiration in the Nordic languages*. Stockholm: Stockholm University. (Doctoral dissertation).
Puggaard-Rode, Rasmus. 2024. Variation in fine phonetic detail can modulate the outcome of sound change: The case of stop gradation and laryngeal contrast implementation in Jutland Danish. *Journal of Phonetics* 106. 101354. DOI: 10.1016/j.wocn.2024.101354.
Raimy, Eric. 2021. Privativity and ternary phonological behavior. In Sabrina Bendjaballah, Ali Tifrit & Laurence Voeltzel (eds.), *Perspectives on Element Theory*, 65–110. Berlin: de Gruyter. DOI: 10.1515/9783110691948-004.
Robinson, Christine & Roibeard Ó Maolalaigh. 2007. The several tongues of a single kingdom: The languages of Scotland, 1314–1707. In Thomas Owen Clancy, Murray Pittock, Ian Brown & Susan Manning (eds.), *The Edinburgh history of Scottish literature*. Vol. 1: *From Columba to the Union (until 1707)*. Edinburgh: Edinburgh University Press.
Russell, Paul. 1988. The Celtic preverb *uss-* and related matters. *Ériu* 39. 95–126.
Salmons, Joseph C. 2020. Germanic laryngeal phonetics and phonology. In Richard B. Page & Michael T. Putnam (eds.), *The Cambridge handbook of Germanic linguistics*, 119–142. Cambridge: Cambridge University Press.
Sammallahti, Pekka. 1998. *The Saami languages: An introduction*. Kárášjohka: Davvi girji.

Sayeed, Ollie & Bert Vaux. 2017. The evolution of Armenian. In Jared Klein, Brian Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*, vol. 2 (Handbücher zur Sprach- und Kommunikationswissenschaft), 1146–1167. Berlin: de Gruyter. DOI: 10.1515/9783110523874-021.

Schrijver, Peter. 2015. Pruners and trainers of the Celtic family tree: The rise and development of Celtic in the light of language contact. In Liam Breatnach, Damian McManus, Ruairí Ó hUiginn & Katharine Simms (eds.), *Proceedings of the XIV International Congress of Celtic Studies, Maynooth 2011*, 191–219. Dublin: Dublin Institute for Advanced Studies.

Ségéral, Philippe & Tobias Scheer. 2008. Positional factors in lenition and fortition. In Joaquim Brandão de Carvalho, Tobias Scheer & Philippe Ségéral (eds.), *Lenition and fortition* (Studies in Generative Grammar 99), 131–172. Berlin: Mouton de Gruyter.

Sidwell, Paul. 2021. Northern Austroasiatic languages of MSEA. In Paul Sidwell & Mathias Jenny (eds.), *The languages and linguistics of Mainland Southeast Asia* (The World of Linguistics 8), 499–546. Berlin: de Gruyter. DOI: 10.1515/9783110558142-023.

Silverman, Daniel. 2003. On the rarity of pre-aspirated stops. *Journal of Linguistics* 39(3). 575–598. DOI: 10.1017/S002222670300210X.

Sims-Williams, Patrick. 1990. Dating the transition to Neo-Brittonic: Phonology and history, 400–600. In Alfred Bammesberger & Alfred Wollmann (eds.), *Britain 400–600: Language and history*, 217–262. Heidelberg: Carl Winter Universitätsverlag.

Sims-Williams, Patrick. 2008. The problem of spirantization and nasalization in Brittonic Celtic. In Alexander Lubotsky, Jos Schaeken & Jeroen Wiedenhof (eds.), *Evidence and counter-evidence: Essays in honour of Fredrik Kortlandt.* Vol. 1: *Balto-Slavic and Indo-European linguistics* (Studies in Slavic and General Linguistics 32), 509–526. Amsterdam: Rodopi.

Sommerfelt, Alf. 1960. The structural point of view applied to the analysis of the consonant system of a Donegal dialect. *Celtica* 5. 107–115.

Spaargaren, Magdalena Jeannette. 2009. *Change in obstruent laryngeal specification in English: Historical and theoretical phonology.* Edinburgh: The University of Edinburgh. (Doctoral dissertation).

Steblin-Kamenskij, M. I. 1974. The Scandinavian consonant shift. *Arkiv för nordisk filologi* 89. 1–29.

Stifter, David. 2012. Lenition of *s* in Gaulish? In Benedicte Nielsen Whitehead, Thomas Olander, Birgit Anette Olsen & Jens Elmegård Rasmussen (eds.), *The sound of Indo-European: Phonetics, phonemics, and morphophonemics*, 523–544. Copenhagen: Museum Tusculanum Press & University of Copenhagen.

Stifter, David. 2015. From Old Irish to Modern Irish. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation*, vol. 3, chap. 114, 2020–2021. Berlin: Mouton de Gruyter.

Stifter, David. 2017. The phonology of Celtic. In Jared Klein, Brian Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*. Vol. 2 (Handbücher zur Sprach- und Kommunikationswissenschaft), 1188–1202. Berlin: de Gruyter.

Stifter, David. 2024. The rise of gemination in Celtic. Version 2. *Open Research Europe* 2023(3). 24. DOI: 10.12688/openreseurope.15400.2.

Ternes, Elmar. 2006. *The phonemic analysis of Scottish Gaelic, based on the dialect of Applecross, Ross-shire*. 3rd edn. Dublin: Dublin Institute for Advanced Studies.

Thomas, Peter Wynn. 1990. The Brythonic consonant shift and the development of consonant mutation. *Bulletin of the Board of Celtic Studies* 37. 1–42.

Thurneysen, Rudolf. 1946. *A grammar of Old Irish*. Trans. from the German by D. A. Binchy & Osborn Bergin. Dublin: Dublin Institute for Advanced Studies.

van Sluis, Paulus, Anders Richardt Jørgensen & Guus Kroonen. 2023. European prehistory between Celtic and Germanic: The Celto-Germanic isoglosses revisited. In Kristian Kristiansen, Guus Kroonen & Eske Willerslev (eds.), *The Indo-European puzzle revisited: Integrating archaeology, genetics and linguistics*, 193–243. Cambridge: Cambridge University Press.

Vaux, Bert. 1998. The laryngeal specifications of fricatives. *Linguistic Inquiry* 29(3). 497–511. DOI: 10.1162/002438998553833.

Vicenik, Chad. 2010. An acoustic study of Georgian stop consonants. *Journal of the International Phonetic Association* 40(1). 59–92. DOI: 10.1017/s0025100309990302.

Villar, Francisco. 1993. Las sibilantes en celtibérico. In Jürgen Untermann & Francisco Villar (eds.), *Lengua y cultura en la Hispania prerromana. Actas del V coloquio sobre lengua y culturas prerromanas de la Península Ibérica (Köln, 25–28 de Noviembre de 1989)* (Acta Salamanticensia. Estudios filológicos 251), 773–818. Salamanca: Ediciones Universidad de Salamanca.

Watkins, Calvert. 1955. The phonemics of Gaulish: The dialect of Narbonensis. *Language* 31(1). 9–19. DOI: 10.2307/410887.

Watson, Seosamh. 2022. *Easter Ross Gaelic*. Dublin: Dublin Institute for Advanced Studies.

Wentworth, Roy Graham. 2005. *Rannsachadh air fòn-eòlas dualchainnt Ghàidhlig Gheàrrloch, Siorrachd Rois*. Dublin: Dublin Institute for Advanced Studies. http://www.celt.dias.ie/publications/online/gearrloch/ (5 June, 2025).

Wheatley, David & Pavel Iosad. In preparation. The fortis/lenis contrast in early 20th century Donegal Irish. MS., University of Edinburgh.

Williams, Nicholas J. A. 2011. Middle and Late Cornish. In Elmar Ternes (ed.), *Brythonic Celtic – Britannisches Keltisch: From Medieval British to Modern Breton* (Münchner Forschungen zur historischen Sprachwissenschaft 11), 237–358. Bremen: Hempen Verlag.

Zec, Draga. 1988. *Sonority constraints on prosodic structure*. Stanford, CA: Stanford University. (Doctoral dissertation).

Chapter 12

Syntactic reconstruction in Celtic

Marieke Meelen
University of Cambridge

Reconstructing syntax has been the most challenging part of research into language history, phylogeny and proto-languages. Problems regarding directionality, transfer and in particular correspondence seemed impossible to address, until recent advances in syntactic theory. In this chapter, I discuss ways to address these issues, focusing on (generative) Minimalist approaches to syntactic reconstruction in particular, with concrete examples from Celtic languages. These methods combine the goal of reconstructing language history, with the analytical tools of formal grammar and cognitive science.

1 Introduction

The comparative historical study of languages, first seen in the work of Marcus Zuerius van Boxhorn and Sir William Jones in the 17th and 18th centuries, laid the foundation for breakthroughs in the reconstruction of a Indo-European proto-language. Historical linguistics textbooks praise the comparative method as the "central" method of reconstruction (Campbell 2013; Trask 1996; Fox 1995; Hock 1991).[1] There are, however, few clear definitions or step-by-step execution guides for the method (Fox 1995: 57; Durie & Ross 1996; Walkden 2014), and it focused

[1]The first visualizations of relations between languages go back to the 17th century with some Germanic varieties by Georg Stiernhielm in 1671 (Sutrop 2012). Other Indo-European languages followed much later (around 1800) in the work by Felix Gallet (Auroux 1990). These early diagrams generally showed networks, however, rather than rooted trees. Trees were popularized in the second half of the 19th century by Auguste Schleicher's 1853 *Stammbaum*, which had 'Celten' as the first ones to branch off from 'Slawogermanen', 'Lateiner', 'Griechen' and 'Arier' (see Schleicher 1853, but also his main work from 1861–1862, as well as Morrison 2016 and Pellard et al. forthcoming for a full history of phylogenetic relations).

Marieke Meelen. 2025. Syntactic reconstruction in Celtic. In Andrew Carnie, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond & Luis Irizarry (eds.), *Foundational approaches to Celtic linguistics*, 417–467. Berlin: Language Science Press. DOI: 10.5281/zenodo.15654879

only on the reconstruction of phonology, morphology and lexical items. This is illustrated by the omission of syntax in Indo-European textbooks (e.g., Beekes 1995; Szemerényi 1996; Krasukhin 2017). Traditionally, Indo-Europeanists have been skeptical about extending this methodology to syntax (Meillet 1954; Jeffers 1976). Because there was no comprehensive syntactic theory in the early heyday of the comparative method, syntactic reconstruction was limited to the placement of individual grammatical items (Wackernagel 1892), or mood and modal categories (Thurneysen 1885 and Delbrück 1893). The fact that publications on PIE alignment systems (e.g., Uhlenbeck 1901; Pedersen 1907 and Vaillant 1936 vs Villar 1983; Rumsey 1987; Lehmann 1993, etc.) and "basic word order" (Dressler 1969; Lehmann 1974; Friedrich 1975; Miller 1975) reached different conclusions only reaffirmed the idea going back to Delbrück himself (Delbrück 1893: v-vi) that "one can no more reconstruct the syntax of a proto-language than one can reconstruct last week's weather, and for the same reason: both reflect chaotic systems" (Lightfoot 2002: 135). Proponents of more recent methods in grammaticalization and syntactic theory (see Section 2), however, have had a more positive outlook on the possibility of unveiling the syntax of earlier languages. In addition to recovering the syntactic history of a proto-language, reconstruction methods can also potentially tell us more about whether and how languages are related and where related languages originate (i.e., the time-lined topology, subgrouping and root of a phylogenetic reconstruction). Hartmann & Walkden (2024) test a number of grammatical factors in a wide range of languages and conclude that syntactic properties are very weak when it comes to encoding phylogenetic information specifically. However, we might also use reconstruction to learn more about how syntax in general or the word order of one specific language can change over time and why it has changed in one particular way (Ferraresi & Goldbach 2008: 9).

In this chapter, I shed light on this last goal of syntactic reconstruction – focusing on Celtic specifically. Despite the fact that the verb-initial word order in insular Celtic languages is an interesting outlier in the family, major publications on Indo-European syntax (e.g., Delbrück 1893; Lehmann 1974) do not discuss any examples from Welsh, Cornish, Breton or Irish. Important early comparative Celtic grammars like Lewis & Pedersen (1961) devote very few words to syntax and are often limited to the position of the verb only (Thurneysen 1927: 301; Lewis & Pedersen 1961; Watkins 1964; Meid 1968). It is not until later that we start seeing comparative syntactic research on the insular Celtic languages. Where work on Celtic syntax initially focused on synchrony (e.g., Harlow 1981; Hendrick 2000; McCloskey 1996; Carnie et al. 2000), in the 1990s, we start to see the first attempts in syntactic reconstruction (e.g., Koch 1991; MacCana 1991; Eska & Evans 1993;

Isaac 1993; Eska 1996 for the verbal complex and basic word order; Willis 1998; 2011 for V2 and free relatives; Newton 2006 for pre-Irish C-domains; Meelen 2016 for British preverbal particles). These studies are extremely useful, but do not yet cover all aspects of Celtic syntactic reconstruction.

I start this chapter by outlining some issues with syntactic reconstruction in general, with specific reference to some examples from Celtic. I discuss how a range of more recent methods have tried to address those problems (Section 2). I then zoom in on generative Minimalist approaches in particular (Section 2.3) in order to show that Celtic languages can teach us a lot about how to develop these methods further (see Sections 3.1 on Celtic and 3.2 on Celtic in contact). Finally, in Section 4, I claim that the reverse is also true, namely that Minimalist syntactic reconstruction can teach us more about the history of Celtic languages.

2 Methods of syntactic reconstruction

2.1 Problems with syntactic reconstruction

Some traditional methods for syntactic reconstruction exploit unidirectional tendencies of grammaticalization. Givón's (1971) adage that "today's morphology is yesterday's syntax" as well as similar work by Gildea (2000b) and Balles (2008), for example, rely on the assumption that morphology exhibits signs of fossilized syntactic items that used to be independent in earlier stages of the language. Similarly, fossilized lexical items whose derivation is transparent could be employed for reconstruction, e.g., present day Welsh *sef* from *ys* '(it) is' +*ef* 'it/this' (Meelen 2016) or *efallai* 'perhaps' from *ef* 'he' and *(g)allai* 'could' (Willis 2011: 14). As Willis (2011) observes, in a VSO language examples like *efallai* that point to SV word order are unmotivated unless in earlier stages of the language SV orders were possible. These studies focusing on individual grammatical items are limited in scope, however, because it is not clear which earlier stage can be reconstructed, nor can it be used for syntactic changes that do not involve grammaticalization.

While sound changes can be constrained by the limits of physical articulators, syntactic *directionality* constraints are harder to discover or measure. A change from unvoiced to voiced consonants between vowels is much more likely than the reverse, for example, but directionality factors of these kind are harder to predict and/or explain in syntax. There are no articulatory restrictions on changes in word order from Object-Verb (OV) to Verb-Object (VO), for example. Typological approaches, such as those by Greenberg (1963), Lehmann (1974), Friedrich (1975), Miller (1975), but also more recent ones like Von Mengden (2008) and Ferraresi &

Goldbach (2008), use implicational relationships between grammatical phenomena to achieve a similar result. Walkden (2009) rightly points out, however, that in order to make any predictions about the syntax, this method requires historical texts in which at least one half of the relevant implicational universal is present. Von Mengden (2008: 109) acknowledges this drawback. For proto-languages, we have no texts by definition and hence no starting point for this method. Walkden (2014: 65) notes, however, that Niyogi & Berwick's (2009: 101–127) model can straightforwardly yield directionality in syntax if it is combined with a cue-based learning algorithm such as that of Lightfoot (1999). As I show later on in Section 4.1, recent discoveries in theoretical syntax – such as parameter hierarchies – may be able to shed more light on this directionality issue as well.

In addition to this so-called "directionality problem", Lightfoot (1980) draws attention to the fact that grammars 'must be created afresh by each new language learner' (Lightfoot 1980: 37), which Willis (2011: 4) has called the "radical reanalysis problem": if children had been able to work out what the earlier structure was, they would not have introduced the new structure. Strictly speaking, this idea of change due to reanalysis in first language acquisition is not unique to syntax. As Harris & Campbell (1995: 372) and Walkden (2009: 23) have pointed out, the same issue arises for phonological reconstruction, where sound change can occur by way of discrete reanalyses (Ohala 1981). Furthermore, Roberts & Roussou (2003), among others, follow the traditional "poverty-of-stimulus" reasoning going back to Chomsky (1980), observing that most of the time, first-language acquisition is remarkably successful despite the lack of direct access to parental grammar. This should follow from Chomsky (2005)'s division of first, second and third factors: "grammars actually acquired on the basis of similar primary linguistic data (PLD) do not vary substantially from one another, otherwise the acquisition task becomes intractable" (Walkden 2014: 49). This extends Willis's (2011: 2) observation that innovative grammars constrain the range of possible hypotheses about earlier stages of the grammar, which thus aids syntactic reconstruction (see Section 2.1 below).

Furthermore, Willis (2011) reminds us that grammatical rules that are present in two or more languages could represent a continuous line of transmission *or* they could be transferred via language contact (the "transfer problem"). In the correspondence sets for phonological reconstruction, borrowings are lexical and much easier to identify and eliminate. Clackson (2017) illustrates this issue with Barðdal's (2013) work on the reconstruction of the construction 'woe is X.DAT' found in a number of Germanic and other Indo-European languages. He argues that the Gothic and Armenian instances of these constructions could also simply have been transferred from Greek through the common bible translations. In

Section 3.2 below, I demonstrate that although this is a serious issue, there are interesting opportunities related to syntactic transfer in reconstruction too.

Another issue that has been raised with syntactic reconstruction is the lack of arbitrariness. Harrison (2003) argues that the real success of the comparative method in terms of making genetic inferences relies on the fact that similarities in symbolic form-meaning pairings (i.e., lexical cognates) "cannot be attributed to nature, and are unlikely to be the result of chance" (Harrison 2003: 223). "Grammatical objects", on the other hand, do not provide any such evidence for genetic relatedness, because form-meaning pairs in syntax are not arbitrary. Where lexical items represent individual linguistic signs that reside in a lexicon, which is "a repository of linguistic unpredictability" (Harrison 2003: 223), grammatical objects are "templates, diagrams, or rules" that are iconic, and not symbolic signs. Iconic signs are complex and do not reflect arbitrary mappings between linguistic categories and the world, which is why they cannot establish genetic relatedness (Harrison 2003: 225). Transferring the comparative method is difficult due to the supposed lack of regularity in syntactic change compared to the famous *Ausnahmslosigkeit* ('exceptionlessness') of sound change (Osthoff & Brugmann 1878; Labov 1981 for the regularity hypothesis in sound change; Lightfoot 2002; Pires & Thomason 2008 for difficulties transferring this to syntax). All the above factors illustrate how difficult it is for syntactic patterns (or "templates, diagrams, or rules") to meet the condition in (1):

(1) Double Cognacy Condition (Walkden 2014: 50)

In order to form a correspondence set, the contexts in which postulated cognate sounds occur must themselves be cognate.

Since any potential higher units of syntax, such as phrases or sentences, are not transmitted from generation to generation in the same way lexical items are, there is no syntactic equivalent to the "cognate" we know from the comparative method of phonological reconstruction (Lightfoot 2002: 125–126). The problem of what to compare instead of cognates, the so-called "correspondence problem" (Watkins 1976: 312), is the most difficult to address. Apart from perhaps some (poetic) stock phrases and idioms, such as PIE *$eg^{wh}ent$ *$og^{wh}im$ 'he killed (the) snake' (Watkins 1995: 301) and *$k̑lewos$ *$n̥d^{h}g^{wh}itom$ 'unperishable glory' (Kuhn 1853: 467), there is no direct correspondence between sentences of one language and another. As syntactic theory has developed in recent years, however, there are now potential answers to search for appropriate alternative comparanda.

In the next section, I discuss some more recent methods of syntactic reconstruction that address this issue. For more detailed exposés on addressing prob-

lems with syntactic reconstruction, see, among others, Bowern (2008), Willis (2011), Walkden (2014), Viti (2015), Clackson (2017), and Gildea et al. (2020).

2.2 Addressing the correspondence problem

Alice Harris and Lyle Campbell suggested that syntactic "patterns" could solve the correspondence problem as patterns can be compared cross-linguistically (Campbell & Harris 2002). They illustrate this with the reconstruction of alignment systems in Kartvelian languages (Harris 2008), using language-internal facts. Similarly, Gildea (2000a) reconstructed the Proto-Cariban verbal system and Kikusawa (2002) focused on alignment systems in Proto-Central-Pacific. While the results of these particular reconstruction efforts may be right, Willis (2011) and Walkden (2014) note a number of issues, such as the lack of psychological reality of "patterns", the unrestricted nature of the method and that "patterns" do not appear to have a clear place in syntactic theory.

This specific issue could be resolved by an approach that uses syntactic patterns stored as units of grammar: Diachronic Construction Grammar (DCxG, see Barðdal & Eythórsson 2012; Smitherman 2015; Eythórsson & Barðdal 2016; Gildea et al. 2020). In DCxG, the "construction" is proposed as the equivalent to the cognate in the comparative method, as they can be compared cross-linguistically (at least partially equivalent to Saussurean arbitrary form-meaning pairs Goldberg 1995, Croft 2001: 18) and, they are hypothesized to be transmitted across generations (Traugott & Trousdale 2013; Barðdal & Gildea 2015). Crucially for syntactic reconstruction, constructions can be either lexical, e.g., *cheeseburger* or schematic, e.g., ditransitive order [SVOO], which adds the semantics of transfer to a sentence like *I'll bake you a pie* (Goldberg 1995). Currie (2013) employs DCxG and in particular the mechanism of "motivation", which, according to him, explains the gradual rise in frequency of Absolute Initial Verb (AIV) order in Early Modern Welsh. When it comes to reconstruction of syntax using DCxG, we find numerous recent attempts such as personal prefixes in Proto-Tupí-Guaraní (Gildea 2002: 322–323), dative experiencers in Germanic (Barðdal & Eythórsson 2012), the subject possessor case marker in the Viceitic (Chibchan) alienable possession construction (Pacchiarotti 2020) and a number of other contributions in Barðdal et al. (2020). Gildea et al. (2020) note that objections regarding constructions fail to meet the Double Cognacy Condition in (1) above (Walkden 2014: 50) – because sentences (i.e., the potential higher order units) they generate are not cognate – are in fact based on a "domain error". They argue that "it is a logical error to require the equivalent of "regular correspondences" when seeking to identify cognates in other domains of language, such as syntax" (Gildea et al.

2020: 18). If this is indeed true we can still call constructions (or "patterns" for that matter) "cognate" simply because they can descend by inheritance from a common ancestor.

Rejecting the need for the Double Cognacy Condition in syntax, however, also means that despite suggested "safeguards" to identify patterns (Harris & Campbell 1995), similarities due to chance, language contact and/or typological factors are hard to identify. Some proposed methods to mitigate these issues, such as avoiding chance similarities (Seržant 2015: 125–130 and Gildea et al. 2020) can only really work in cases of identity or strong similarity of the proposed cognates in the daughter languages. This, Willis (2011) observes, is a general weakness of the CxG approach to syntactic reconstruction: when there is structural reanalysis (and thus a potential lack of identity in one or more daughter languages), this type of reconstruction can only work if it can rely on a straightforward account of how constructions are reanalyzed, innovated or dissolved. Traugott & Trousdale (2013: 35–38)'s "neoanalysis" and "analogization" provide ideas in this direction. When it comes to identifying similarities due to language contact, Daniels (2017) acknowledges this is an important issue and aims to provide a method for doing exactly this. His method exploits the possibility of a third type of construction, which is "somewhere between lexical and semantic" (Daniels 2017: 382), e.g., as in [the X-er the Y-er] (*The more you read, the less you understand*) (Fillmore et al. 1988: 506–508). Daniels (2017) builds on earlier work by Ross (2015) and Seržant (2015), suggesting that in order to exclude contact, after hypothesizing correspondence sets, it is necessary to follow the next step in the traditional comparative method ensuring that the phonological material represented in the partially schematic constructions is in fact inherited. This then could work for those constructions that have at least some inherited lexical element, but it is hard to see how this method could be extended to purely schematic constructions. Even those with lexical elements cannot always provide a straightforward answer, as Clackson (2017) has shown with his evaluation of the *woe*-DAT construction. Finally, the lack of reference to a theory of language acquisition makes it harder to extend the reconstruction of similar constructions to the reconstruction of a grammar that reaches "the same level of analysis that one would expect for an attested language" (Willis 2011: 10). In the next section, I discuss recent generative approaches to syntactic reconstruction that have aimed to do exactly that.

2.3 Minimalist approaches to reconstruction

Using the Minimalist Program, Van Gelderen (2011) formalizes syntactic change through a feature cycle in which semantic features (e.g., a first-person pronoun) turn into interpretable formal features [iF:1s] and, finally, into uninterpretable features [uF] that can be probes. She connects common cases of syntactic reanalysis to labeling of phrases, which is seen as a driving factor of change (Van Gelderen 2022). Using grammatical (formal) features in syntactic comparison and change goes back to the Borer-Chomsky Conjecture (Baker 2008: 155), which states the following shown in (2):

(2) Borer-Chomsky Conjecture (BCC)

All parameters of variation are attributable to the features of the functional heads in the lexicon.

Walkden (2009, 2014) observes that this offers a solution to the correspondence problem in cases where functional items are overt. Cognacy of lexical items can be established in the usual way employing all the advantages of the traditional comparative method for reconstruction, which means there is little room for chance or contact-induced similarities.

Willis (2011) uses the example of British Celtic free relatives, i.e., the cognates of Modern Welsh *bynnag* '-ever', Middle Welsh *bynnac*, Middle Breton *pennac* and Middle Cornish *penagh*. Just like in English, all these languages allowed for those forms to follow a *wh*-word (cf. English *whoever*, *whatever*, etc.), but they differ in whether they allow for both strong and weak *wh*-forms preceding reinforcing elements, e.g., Middle Cornish *penag-**el*** 'who-ever-**all**', etc. He first describes the exact similarities and differences between the three languages and then asks what grammar we should reconstruct for the proto-language, before carefully laying out every stage of the abductive reanalyses and extensions. He starts from the output of Common British *py nag el* 'whoever goes' with the pronoun *py* with an interpretable *wh*-feature in SpecCP and the negator *nag* + the verb *el* incorporated into the C-head. At the next stage, the negator joins the *wh*-pronoun in SpecCP opening up a slot in the C-head which can then, after extension, be filled with the regular pre-verbal particle *a*, resulting in Middle Welsh *bynnag a el* 'whoever PCL goes' (see Willis 2011: 25–36 for a full analysis). The strength of Willis's reconstruction lies in the meticulous attention to detail in justifying every reconstructed stage and form. He not only tests to what extent we can use traditional techniques for reconstruction such as majority and economy, but also refines the notion of directionality distinguishing "local directionality"

from "universal directionality", thus imposing important constraints on the possible reconstructions: the first testing plausible reanalyses that are consistent with detailed aspects of the known daughter language systems, the second through "broad-brush rules of thumb" akin to that in phonological reconstruction (Willis 2011: 36). Finally, through describing and analyzing the microvariation between the daughter languages and carefully checking each stage of reanalyses and extensions, he tests which outputs could result from each stage of the reconstructed grammars, thus addressing the "radical reanalysis" problem. This method may be rather limited as it can generally only be applied to overt lexical items with functional features that have cognates in other languages, which means it also cannot work well for establishing reconstructed stages of languages that go very far back in time. However, the results are extremely reliable when each stage of the previous grammar and the various reanalyses and extensions are carefully checked. Further examples of exploiting true cognates with functional features can be found in Walkden (2009, 2014) for Germanic and Meelen (2017) for preverbal particles in British Celtic.

Another recent Minimalist attempt at syntactic reconstruction uses the Parametric Comparison Method (PCM) pioneered by Guardiano & Longobardi (2005) but extended significantly based on Roberts (2019)'s approach to microvariation and parameter hierarchies. The Parametric Comparison Method combines the goal of reconstructing language history with the analytical tools of formal grammar and cognitive science. This method ultimately goes back to the 1980s, with work on syntactic Principles and Parameters (P&P) of Universal Grammar (UG), which concerns language acquisition, language change, markedness, implicational relations, learnability theory and opportunities for cross-linguistic comparison. Chomsky (1980) draws on Jacob's (1977) ideas on biological diversification making a specific analogy to linguistics: "the problem of accounting for the growth of different languages... is not unlike the general problem of growth, or for that matter, speciation" (Chomsky 1980: 67). The P&P approach offered solutions for Plato's Problem (i.e., the poverty of stimulus in language acquisition) and made excellent predictions for typological variation. However, within a Minimalist approach to cross-linguistic variation, the idea that parameters are part of the genetic endowment, the innate UG, is difficult to maintain, which is exactly what Roberts (2019) aims to remedy since proposals that abandon the notion of parameters altogether cannot offer viable alternatives (for a full discussion, see Newmeyer 2005 and reviews and replies by Dryer 2007 and Walkden 2014, Berwick & Chomsky 2011, Boeckx 2015, Roberts 2019: 11–13).

Ledgeway & Roberts (2017) follow Chomsky (1995: 6) in the Borer-Chomsky Conjecture (BCC): all microparameters are properties of individual heads,

i.e., the formal features of functional heads in the syntactic structure. Cross-linguistic variation in syntax is variation in the values of these features (e.g., the C(omplementizer) head having C[+wh] vs C[−wh] to capture the distinction between *wh*-movement and *wh*-in-situ languages). This aligns Roberts' earlier work (Clark & Roberts 1993, Roberts 1997, Roberts & Roussou 2003, Roberts 2007) with Chomsky's (2005) proposal of the three factors of language design and makes it more concrete:

(3) Three factors relevant for parameters (Roberts 2019: 7)
   a. Factor 1: Underspecification of formal features in UG
   b. Factor 2: Trigger experience/what the child takes in (PLD)
   c. Factor 3: Feature economy (FE) and input generalization (IG)

Unlike earlier parametric theories, formal features are therefore not specified by UG, and parametric variation arises from the interaction of the three factors. Biberauer & Walkden (2015) and Biberauer (2019) add that children postulate formal features on functional heads in cases where the PLD intake departs from the simplest Saussurean form-meaning mapping such as cases of discontinuity and displacement (traditional "Move" or "Internal Merge"), multiple realization of elements (e.g., due to morphosyntactic agreement) and apparent non-realization of meaningful elements (i.e., silence). Similarly, we can recover a parameter setting for a particular language by looking at the overt realization of features, e.g., a participant, gender and/or number feature could be overtly realized through inflection or a clitic, as shown in the Fiorentino inflectional paradigm in (4):

(4) (e) parlo, tu parli, e/la parla, si parla, vu parlate, e/la parlano
'I/you/(s)he/we/you/they speak(s)' (Fiorentino; Brandi & Cordin 1989)

In addition to this straightforward spell-out (exponence, i.e., equivalent to F ↔ /p/ in Distributed Morphology), features could be "realized" in a more indirect way, e.g., when feature values are copied into functional heads via Agree with a defective goal leading to head-movement (Roberts 2010). When features are spelled out overtly, this approach comes close to the approach adopted by Walkden (2009, 2014) and Willis (2011) above, with the distinction that feature exponents in the microparametric approach do not necessarily meet the Double Cognacy Condition, unless their exponents are in fact "true lexical cognates" that can be independently verified as "inherited" through their sound correspondences and the Regularity Hypothesis. The main advantage of reconstruction based on "true lexical cognates" resulting from the Double Cognacy Condition is that it allows us to eliminate resemblances due to chance, contact or typological factors. Examining cross-linguistic variation based on one single microparameter

would therefore yield rather limited results, which will need to be independently checked very carefully in order to establish whether they represent an inherited feature or not. Chance resemblance is furthermore problematic as all parameter settings are binary and "horizontal" transfer due to contact may be harder to detect in syntax than it is for lexical items. The Parametric Comparison Method for examining cross-linguistic syntactic variation therefore relies on not one, but many parameters, which effectively nullifies the opportunity to find chance resemblances of every parameter setting between any pair of languages and also reduces the potential impact of contact factors (see Section 3.2 for how the PCM could actually exploit potential lack of contact factors too). Typological factors, finally, are actually built into the method in the form of implicational factors and parameter hierarchies to which I turn now.

Baker (2008) observed that on the microparametric view where variation lies in the feature make up of functional heads (i.e., the BCC) "there should be many mixed languages of different kinds, and relatively few pure languages of one kind or the other" (Baker 2008: 350). Macroparameters (Baker 1996; 2008; Bošković 2008; Huang 2015), on the other hand, were traditionally proposed to be directly associated with UG principles, e.g., the Polysynthesis Parameter (Baker 1996), which is a general notion of "argument visibility" with two possible options: syntactic configurations vs formation of complex words. A macroparametric view on cross-linguistic variation predicts that there is a rigid division of all languages into clear types (e.g., OV vs VO, polysynthesis or not, etc.). This means that every category in every language should pattern in one way or another. The empirical evidence, however, shows that instead we have a more bimodal distribution: when it comes to word order, for example, we find many languages clustering around one type or another, but there are also some exceptions to these principal patterns. This suggests that we need a combination of both micro- and macroparameters to adequately account for cross-linguistic variation. Based on Kayne (2005: 10) among others, Roberts (2019) argues that in addition to formal feature values of functional heads, there are Independent Factors (IF) "which can cause groups of heads – over which we can generalize with an appropriately formulated feature system – to act in concert" (Roberts 2019: 5). This essentially turns macroparameters into a surface effect of aggregates of microparameters acting in unison, creating a strict hierarchy of parameters, shown in Figure 1.

Biberauer & Roberts (2012, 2017) thus propose a typology of parameters where macroparameters share a value of a variant feature on ALL heads; mesoparameters share it on ALL heads *of a given class* (e.g., [+V]), whereas microparameters show that value on only a small, lexically-definable subclass of functional heads (e.g., subject clitics or modal auxiliaries) and, finally, nanoparameters have this

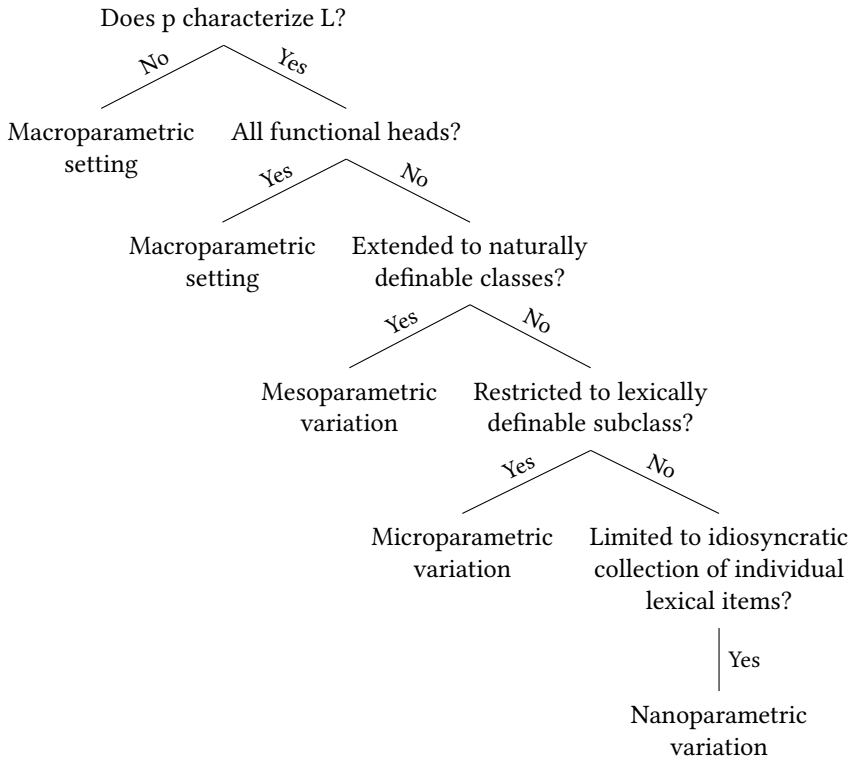


Figure 1: Parameter Hierarchies from Ledgeway & Roberts (2017: 595)

value specified to just one or more individual lexical items. This hierarchy makes specific predictions for syntactic typology, but also for language acquisition and change. Typologically, macroparameters are expected to be pervasive or universal, affecting a large number of all relevant functional heads in the grammar of a language. Macroparameters should be salient in the Primary Linguistic Data (PLD) and therefore easier to acquire and diachronically thus be more stable than microparameters. Although this is still work-in-progress, a number of hierarchies have already emerged from the work by the Cambridge ReCoS group and others based on empirical evidence from a wide range of languages (see Biberauer et al. 2014, Sheehan 2014, Van der Wal 2020, Ledgeway 2012, Ledgeway & Roberts 2017, Kinn 2017, Douglas 2017 and mainly Roberts 2019 and references therein).[2]

[2]See also https://recos-dtal.mmll.cam.ac.uk/Publications/Published.

This radical revolution in parametric theory, with well-established data-driven parameter hierarchies, gives the Parametric Comparison Method (PCM) a very solid ground for the reconstruction of language history. The PCM employs a parameter grid with binary values (+/–) for each discrete parameter in every language under investigation. This is based on a growing list of implicationally organized parameter values consisting (in its most recent iteration) of 94 nominal parameters tested on 69 languages (Longobardi 2017; Ceolin et al. 2020), as well as at least 87 verbal parameters for 39 languages (Roberts et al. 2022; Baker & Roberts forthcoming). These hierarchically organized implicational parameters differ significantly from earlier attempts to use a list of binary features to encode "language structure" such as Nichols (1992), Ringe et al. (2002), Dunn et al. (2005). The latter only look at straightforward surface patterns without taking into account what kind of grammar could generate these patterns. Similarities in surface patterns are interesting but do not necessarily indicate that languages are historically closely related, even if the pattern in question is typologically unusual. Jongeling (2000) draws our attention to this listing numerous surface patterns in Welsh that look remarkably similar to Biblical Hebrew and finds an even closer parallel in Phoenician, for example, with the unusual Determiner-Noun-Demonstrative order shown in (5):

(5) a. y llances hon [Welsh]
the girl this
'this girl'

b. ha- nnaᶜərā ha- zōt [Biblical Hebrew]
the girl the this
'this girl'

c. h gbr z'ɔ [Phoenician]
the man this
'this man' (Jongeling 2000: 81)

Similarly, Roberts (1998) notes that in Greek, for example, the same linear and concord pattern (Noun-Genitive-Adjective) also exists in Hebrew, but Longobardi et al. (2013) demonstrate that these similar surface patterns in fact reflect opposite values in two distinct parameters. In essence, the PCM here therefore requires the same amount of in-depth grammatical analysis as that shown by Willis's (2011) reconstruction of the various stages of free relatives in British: the question of how the surface patterns can be generated remains of crucial importance.

A simple example of an implicational relation in the parameters is shown by Longobardi & Guardiano (2009) as Nominal Parameter 23 (Pn23), which distinguishes languages that exhibit concord (at least) in Number with a subject nominal in predicative constructions. This parameter is contingent on the positive setting of two other parameters, namely whether there is a Number feature on N (i.e., Parameter 5 should be positive), which in turn can only be positive if there is a Number feature in the language in general (i.e., Parameter 2 should be positive). Longobardi & Guardiano (2009) note in their extensive Appendix listing all parameters in detail that this is the parameter that separates languages like English, German, Irish, Welsh and Farsi with uninflected predicative adjectives from the rest of Indo-European, as shown by examples of English and Irish, on the one hand (shown in 6), and Standard Italian and Triestino[3] (Northern Italian variety spoken in Trieste), on the other (shown in 7 and 8):

(6) a. Tá an fear beag. [masc. predicative]
be.3s the man small
'The man is small.'

b. Tá an bhean beag. [fem. predicative]
be.3s the man small
'The woman is small.'

c. Tá na daoine beag. [pl. predicative]
be.3s the people small
'The people are small.'

d. Na daoine beag**a**. [pl. attributive]
the people small.PL
'the small people'

(7) a. L'uomo è piccol**o**. [masc. predicative]
the.man be.3s small.SM
'The man is small.'

b. La donna è piccol**a**. [fem. predicative]
the woman be.3s small.SF
'The woman is small.'

c. Le persone sone piccol**e**. [pl. predicative]
the people are small.PL
'The people are small.'

[3]Many thanks to Christian Faggionato for these examples.

d. le persone piccol**e**. [pl. attributive]
the people small.PL
'the small people'

(8) a. L'omo xe pici**o**. [masc. predicative]
the.man is small.SM
'The man is small.'

b. La dona xe pici**a**. [fem. predicative]
the woman is small.SF
'The woman is small.'

c. Le persone xe pici**e**. [pl. predicative]
the people are small.PL
'The people are small.'

d. le persone pici**e** [pl. attributive]
the people small.PL
'the small people'

The PCM can be used to compare a small number of individual parameters (or even a single parameter) related to one particular grammatical feature, as is the case of Number in the DP with parameters 2, 5 and 23 above. These can also be compared in different diachronic stages of a single language to trace syntactic changes over time (see Section 4.1 below).

However, another potential strength of the PCM lies in its ability to compare a large number of parameters in many different languages to create a phylogeny that can tell us how languages are historically related to each other. One advantage is, for example, that this method can be applied to any set of languages, regardless of how similar or different they are, because the parameters are drawn from a finite, universal list. Their discrete, binary nature lends itself to precise calculations of differences between languages. There are two main methods for these types of phylogenetic computational comparisons: distance-based methods (e.g., Neighbor joining, FITCH/KITCH, UPGMA) or character-based methods (e.g., Maximum Parsimony, Bayesian analysis). The latter type of methods are based on optimizing the probability of a proposed tree given the observed data, which mainly consists of databases of lexical cognates such as Dyen et al. (1992). These methods have gained much popularity in computational historical linguistics in the last decades and are mainly used to show taxonomies as well as depth of splits of family branches (Holden 2002; Atkinson & Gray 2005; Bouckaert et al.

2012, etc.). Character-based methods crucially rely on the assumption that characters (in this case lexical cognates) are independent. The arbitrary form-meaning pairing of cognates makes them plausible candidates for this type of probabilistic modeling. However, as Jäger (2013), among others, has observed, this level of independence cannot be maintained when it comes to the regular sound changes: "It rests on the simplifying assumptions that mutations at different positions are stochastically independent and that mutation probabilities are constant across lineages" (Jäger 2013: 286.) More recent character-based studies aim to address earlier issues with this method in various ways, e.g., through finding methods to create better and more reliable input data (especially for lesser-studied language families where extensive cognate databases are still a desideratum, see List et al. 2016; 2017) or by adding specific priors (e.g., Rama 2018), weights or other additional information (e.g., Neureiter et al. 2022 attempting to include the possibility of horizontal transfer, i.e., language contact). Although these methods have achieved interesting results, with a precise representation of the history of a language family including potential dates of splits and information about how reliable each part of a given taxonomy is, Hammarström et al. (2019) and others have pointed out that some issues remain with forcing cognate classes to be binary characters (as is necessary for these models). Lexical cognates are not always perfect pairs of 1 form + 1 meaning, and often meanings have changed to the extent it is hard to establish cognacy. Finally, certain meanings (or lexical items) can be lost from the language entirely (Meelen et al. 2022).

Parameters, on the other hand, are binary by definition, but since they interact and are part of implicational hierarchies, they cannot form the basis of off-the-shelf methods as they would violate the independence of character assumption. The effect of these interdependent relations between parameters is not negligible: compared to total independence, the number of possible languages is reduced by seven orders of magnitude (with a set of 75 parameters) showing that implicational relationships between parameters play a key role (see Bortolussi et al. 2011 and Longobardi 2017 for tests of statistical significance of this and other parts of the PCM). Similarly, Ceolin et al. (2021) calculate that nearly 40% of the taxonomic input is neutralized in a dataset of 94 nominal parameters set for 58 languages. For this and other reasons (see McMahon & McMahon 2005 and others), the PCM uses distance-based methods to create phylogenies instead. As opposed to character-based algorithms, distance-based ones use pairwise distances between languages based on the identities and differences in the comparanda, i.e., each of the parameters set positively "+" or negatively "–". In addition to these binary "±" values, parameter settings could be redundant or irrelevant indicated by '0' as they are implied by a particular setting of other parameters as

we have seen in the interdependent relation of the parameters referring to the Number feature above. Finally, the positive setting of a parameter requires some specified empirical evidence (Crisma et al. 2020 and the discussion on the realization of features above), whereas negative settings represent default values. It follows then that neither pairs that contain '0', nor pairs that are both default (Language A "–" *and* Language B "–"), should be part of the distance calculation, which is exactly what the Jaccard distance, shown in (9), corresponds to (with N as the number of pairs of parameter settings in two languages A and B):[4]

(9) $$\Delta\text{Jaccard}(\text{Lang}_A, \text{Lang}_B) = \frac{[N_{-+} + N_{+-}]}{[N_{-+} + N_{+-} + N_{++}]}$$

Ceolin et al. (2020, 2021), Longobardi & Treves (2023) and others demonstrate further applications of the PCM ranging from the possibilities for the reconstruction of "deep language history" to neural implementations, i.e., how can binary parameters be conceivably implemented in the cortical circuitry in the human brain? However interesting these opportunities, I leave those aside in this chapter and instead, before moving on to more Celtic data, highlight one important observation that has direct implications for the syntactic history of a set of languages, namely, the "Anti-Babelic principle":

(10) Anti-Babelic principle (Guardiano & Longobardi 2005)

Similarities among languages can be due either to historical causes (common origin or, at least, secondary convergence) or to chance; differences can only be due to chance (no one ever made language diverge on purpose)

Although the strict version of the Anti-Babelic principle cannot hold (see, among others, examples in Thomason 2007), deliberate language change is certainly not the norm and there is ample evidence that syntactic features in particular, unlike phonetic features, are less likely to be consciously adjusted (see, for example, Smith 2000). The importance of this principle for the PCM is that it makes clear predictions about historically significant distributions of languages. If languages are known to be closely related, we expect a distance δ between 0 and 0.5, whereas the distances of other languages should tend towards 0.5 and there

[4]Note that in earlier presentations of the PCM, e.g., Longobardi & Guardiano (2009), Normalized Hamming distances were used leaving out the potential significance of negative settings as the default. As work on the PCM develops and more research goes into the exact parameters and their hierarchical relations, it is clear that Jaccard distances yield a better representation of the historical linguistic situation. See supplementary materials in Ceolin et al. (2021) for further discussion.

should be very few (if any) language pairs with a distance between 0.5 and 1. Baker & Roberts (forthcoming) decide, based on this and calculations by Bortolussi et al. (2011), that any value of $\delta < 0.20$ indicates probable genetic relatedness, whereas any value of $\delta > 0.40$ is likely to be random. They illustrate this with results from 87 verbal parameters in known language families where Celtic, Slavic language family and Romance languages have an average $\delta = 0.15$ and Sinitic and Semitic have an average $\delta = 0.18$ and $\delta = 0.19$, respectively. Larger families such as Germanic or all of the Indo-European languages in their sample ($n = 39$) have $\delta = 0.22$ and $\delta = 0.27$ respectively, whereas a comparison of all languages in their database yields $\delta = 0.37$. By comparison, if we focus on just 56 parameters of the nominal domain, Celtic languages are even more closely related with $\delta = 0.02$ (according to Longobardi 2017: 257).

In the next few sections, I first show why Celtic languages are so important for the development of these types of Minimalist syntactic reconstruction methods. Since there already are good examples in Celtic of using cognate functional items such as Willis (2011)'s reconstruction of British free relatives summarized, I focus here on parametric Minimalist methods only. For both the verbal and the nominal parametric distance results discussed above, only data from Modern Irish and Modern Welsh were taken into account, but here I extend this with data from older stages of these languages (Middle Welsh and) as well as occasional examples from other Celtic languages like Cornish and Breton. I then show how the PCM in particular can help advance research on Celtic languages. Finally, I discuss some strengths and limitations of both types of Minimalist approach, ultimately arguing that we need both for a more comprehensive syntactic reconstruction.

3 How Celtic can help Minimalist syntactic reconstruction

The Celtic branch of Indo-European forms an ideal testing ground for the PCM. First of all, it contains a number of languages, both with attested older stages and with languages that are still spoken in the present day. The latter gives us the opportunity to get native-speaker judgments on every possible parameter providing the kind of detailed syntactic information that is necessary for in-depth analyses. As for the older stages, Celtic has both inscriptions and manuscripts from a wide range of times and places. These show that while there is no doubt that languages like Lepontic, Gaulish, Irish and Welsh are related, they exhibit numerous differences – especially in the syntax of their oldest stages. Whereas some branches of Indo-European are relatively uniform in terms of, for example,

"basic word order" (e.g., verb-second in Germanic), Continental Celtic inscriptions show important variation from their Insular Celtic counterparts. These differences within a set of languages that have been grouped together through the traditional comparative method of phonological reconstruction provide us with the opportunity to enhance and refine the set of parametric questions we ask. Finally, from a purely geographical point of view the Celtic languages are of crucial importance in the reconstruction of Indo-European as they provide an excellent test case for what can happen at the forefront of Indo-European expansion (Ram-Prasad & Meelen 2022). In what follows, I show how certain aspects of Celtic syntax can help to refine the set of parameters used in the PCM, and how evidence from some Celtic languages like Middle Cornish can provide more insight into issues of syntactic transfer in contact situations.

3.1 Celtic languages refining the PCM

From a syntactic point of view, the Insular Celtic languages in particular are interesting since their verb-initial nature sets them apart from other Indo-European languages. Other syntactic features that are found in Celtic, but not in other Western European languages include inflected prepositions, shown in (11) with examples from Cornish and Breton, and the lack of a 'have'-like predicate to express possession in Welsh and Irish in particular, shown in (12).)

(11) warnav/-nas/-nodho/-nydhi/-nan/-nowgh/-nedhe [Middle Cornish]
warnon/-nout/-nañ/-ni/-nomp/-noc'h/-nezho [Breton]
'on me/you/him/her/us/you(PL)/them' (Toorians 2014 & Wmffre 1998: 30)

(12) a. Mae gen i feic/annwyd [Welsh]
is with.1s me bike/cold
'I have a bike/cold.'

b. Tá teach ag an mbean. [Irish]
is house with the woman
'The woman has a house.'

Modern Breton, however, does have a lexical verb 'have' namely *kaout/endevout* (shown in 13a), but this verb is historically derived from the verb *bezañ* 'be' + a prefixed personal pronoun. Breton can also indicate possession using a preposition *gant* 'with' (shown in 13b), which is cognate with Welsh *gan* 'with' above.

(13) a. Bremañ Azenor ha Iona o deus un ti [Breton]
now Azenor and Iona 3P have.PRES a house
'Azenor and Iona have a house now.' (Jouitteau 2005: 373)

b. Ganti e oa teir yar. [Breton]
with.3SF PRT be.IMPF.3S three hen
'She had three hens.' (Press 1986: 140)

However, these examples of syntactic variation within and beyond the Celtic language family are not covered by the currently-available sets of parameters in the nominal and the verbal domains (see Crisma et al. 2020 for 94 nominal parameters and Baker & Roberts forthcoming for 87 verbal parameters).

Furthermore, although insular Celtic languages show remarkable similarities in terms of syntax, especially in the nominal domain, there are also cases of (micro)variation. Another example of such variation that is currently not captured by the PCM framework concerns resumptive pronouns. Resumptive pronouns are generally not available in subject positions cross-linguistically, even in languages like Irish which usually allow resumptive pronouns in other positions, e.g., direct objects, shown in (14):

(14) a. an fear a raibh (*sé) breoite [Irish subject relative]
the man rel was he ill
Intended: 'the man who (he) was ill' (McCloskey 1990: 210)
b. an fear ar bhuail tú é [Irish object relative]
the man rel struck you him
'the man that you struck (him)' (McCloskey 1990: 206)

In Modern Irish, resumptives are banned in subject position only in the main clause of a *wh*-construction. This so-called "Highest Subject Restriction" is also found in Hebrew and Palestinian Arabic (Shlonsky 1992 and Alexopoulou 2006). In Modern Welsh, on the other hand, resumptive pronouns are not allowed in subject position in main clauses and in those embedded clauses where a gap is possible (Borsley et al. 2007: 108).

(15) a. y dyn gafodd (*e) 'r wobr [Welsh subject relative]
the man get.PAST.3S (he) the prize
'the man who got the prize'
b. y ffrwydrad glywais i (*e) wedyn [Welsh object relative]
the explosion hear.PAST.1S I (it) then
'the explosion I heard then' (Borsley et al. 2007: 119)

Welsh does allow resumptives in relatives based on possessives and objects of prepositions, as shown in (16):

(16) a. y myfyrwyr werthodd Ieuan y ceffyl iddyn **nhw** [Welsh]
the students sell.PAST.3S Ieuan the horse to.3P them
'the students who Ieuan sold the horse to'

b. y dyn welais i **ei** chwaer [Welsh]
the man see.PAST.1S I 3SM sister
'the man whose sister I saw' (Borsley et al. 2007: 121)

This Welsh pattern is also found in other languages, e.g., Czech (Toman 1998) and Berber and Turkish (Ouhalla 1993). McCloskey (1990) and Willis (2000), among others, have analyzed this as a constraint on binding with respect to pronouns in certain domains (the A'-Disjointness Requirement). More research on this microvariation in the distribution of resumptive pronouns in languages like Irish and Welsh can therefore clearly provide more comparanda for the PCM, which in turn will help it make better representations of differences and similarities between subbranches in particular, but also across language families.

When it comes to the nominal domain, the first set of large-scale parametric comparisons to create phylogenetic trees (Guardiano & Longobardi 2005) was based only on a set of 54 distinct parameters. As research on the PCM evolved, these parameters became more and more refined and further parametric distinctions were added. This was partly due to new discoveries in the field of nominal syntax, but also because more languages, especially non-Indo-European ones, were added to the PCM databases. Even within the most recent publications using a total of 94 nominal parameters and a total number of 69 languages from a wide range of language families, Celtic is only represented by Modern Irish and Welsh, however. It is therefore clear that more research on the syntax of Celtic languages, especially older stages as well as those not yet queried by the PCM, such as Scottish Gaelic and Modern Breton, can help improve Minimalist syntactic reconstruction.

3.2 Celtic languages in contact

It is easier to detect similarities due to language contact with Neogrammarian lexical cognates. There is no clear equivalent to the regularity of sound laws in syntactic change, but parameter hierarchies and the implicational relations found in theoretical syntax do provide some opportunities to distinguish inherited structure from transfer due to language contact. Ledgeway & Vincent (2025), for example, show that syntactic changes due to contact do not follow the step-by-step parameter hierarchy from macro- to meso-, micro- and nano- or vice versa. Recall from Section 2.3 that nanoparameters involve features on individual lexical

items, whereas microparameters refer to a lexically definable subclass of functional heads, e.g., auxiliary verbs. Mesoparameters deal with core functional categories or larger natural classes of functional heads and their effects. They differ from macroparameters as they involve formal features of some but not *all* functional heads acting together, such as head-movement in Germanic and Romance. When it comes to endogenous changes, we expect parameters to change step-by-step from one level to the next, as shown by Romance active participle agreement. Ledgeway & Vincent (2025) observe that parametric comparison of Romance and Greek varieties spoken in the south of Italy instead show abrupt and saltatory movements through parametric space indicating exogenous change, i.e., changes due to contact. The Romance language Aspromonte Calabrese, for example, which has been in contact with Greko (a variety of Greek spoken in Italy), has lost auxiliary selection completely, "jumping" from a nanoparametric setting with auxiliary selection sensitive to argument structure to a macroparametric setting found in Greko that has no auxiliary selection at all. Similarly, in Old Calabrese syncretic marking of internal arguments was restricted in a nanoparametric fashion to 1/2 person only. Endogenous changes in this and non-contact varieties in Romance led to the loss of this 1/2 person restriction, moving one step up the hierarchy. However, contact with Greko then led to a subsequent jump up two levels much closer (but not identical) to the level of Greko, showing that exogenous changes can lead to hybridism which is more like approximation than convergence, but still predicted through the strict distinctions made by parameters. Although the biggest leap or "catapult effect" from nano- to macroparametric setting is actually predicted for some forms on endogenous change as well (e.g., passives), Ledgeway & Vincent (2025) show that parameter hierarchies can still be used as a good heuristic for detecting contact-induced change.

We can explore the issue of transfer further through investigating a group of closely-related languages, where we know some have undergone more contact-induced change than others. The British Celtic languages form an excellent case study, since both Breton and Cornish are clearly heavily influenced by French and English respectively, with which they were/are in close contact. For Cornish, Padel (2021) notes that "Cornish shows extensive influence from English" (256), even in the earliest texts going back to the 15th century, because Cornwall by that time had been an administrative county of England for five centuries already. Further evidence comes from the widespread use of Anglo-Saxon forenames in the area as well as the numerous loanwords that are fully assimilated into Cornish found in the *Old Cornish Vocabulary*, dating from the later twelfth century (Graves 1962). Humphreys (1992) finds similar patterns for Breton, which

is heavily influenced by French already in the medieval period as it ceased to be the language of the ruling class in the 12th century.

When it comes to syntax, there are clear differences between Breton and Cornish, on the one hand – sharing certain features with French and English, respectively – and Welsh, on the other. This poses the question of whether these differences are due to language contact. Where Irish and Welsh lack a lexical verb 'have', as shown in Section 3.1 above, both Breton and Cornish have innovated these through the grammaticalization of an infixed object pronoun and a form of the verb 'be', as shown in (17a) for Middle and (17b) for Modern Breton:

(17) a. mil uileny diouz tut an bro ouz bezo huy
1000 insult from people the land 2PL be.FUT.3S you
'you'll have 1000 insults from the people of the land' (B 176.766)

b. Ur c'hi ac'h eus? - N 'em eus ki ebet.
Q dog have.2S - NEG have.1S dog none
'Do you have a dog? - I haven't got a dog.' (Press & Ar Bihan 2004: 57)

Although this grammaticalization could potentially be reconstructed for Proto-British if occasional instances in early poetry show this innovation was incipient in Welsh as well (Lewis & Piette 1966: 50–52), the increased level of contact with languages that did have lexical 'have' like English and French could have accelerated this grammaticalization in Breton and Cornish. The subsequent developments of using the verb 'have' to form a perfect is directly modeled on French (for Breton, cf. Willis forthcoming). For Cornish, Williams (2011: 328) cites one example shown in (18), which could be influenced by Middle English. However, Padel (2021: 258) argues that Cornish never developed an analytic perfect with the verb 'to have', and (18) shows a stative verb, so this subsequent change may only have occurred in Breton.

(18) flehys a m bes denethys (Middle Cornish)
children PRT 1S have.3S beget.PPP
'I have begotten children' (CW 1979)

Another example of grammaticalization is the development of the Cornish volitional verb *mynnes* 'to wish' to a future marker. This is crosslinguistically common (e.g., Danish *ville*, Greek *tha* < *thelo ina* 'I wish that', Bulgarian *šte* < 'I want that'), but not found in Welsh or Breton. Old English *wile* 'want, wish' (> PDE future *will*) unlike other PDE auxiliaries that only grammaticalized later, could already have future meaning before the Middle English period (Warner 1993).

This early grammaticalization in English makes it a good candidate for contact-induced, or at least contact-accelerated grammaticalization in Cornish.

Similarly, if we compare the word order in declarative matrix clauses across the medieval British languages, we can see a clear V2 pattern, which can be reconstructed to Proto-British (alongside V1) (Meelen 2016, Willis forthcoming). In Middle Cornish, however, there are many more exceptions to strict V2 with subject pronouns following initial constituents, shown by the following examples from Eska & Bruch (2021: 151, 155):

(19) a. an dragan me a ra guan [Middle Cornish]
the dragon I PRT make.PRS.3S weak
'I shall make the dragon weak.' (BM 232.4004)

b. the 'th scoforn wharre yehes sur me a re
to your ear soon health surely I PRT give.PRS.3S
I shall give, surely, healing to your ear soon.' (PC 312.1150–1151)

These could be due to poetic meter (Eska & Bruch 2021). However, the examples with unstressed subject pronouns are also found in neighboring Old English and southern Middle English (Kroch et al. 2000), language contact is very likely to have influenced this Cornish innovation as well (Willis forthcoming).

An example of an innovation in Middle Breton and Cornish that is very clearly modeled on French and English, respectively, is the passive. Similar to Irish, voice alternations in Brittonic can be expressed through impersonal verbal morphology, which is present across all three medieval languages, e.g., Middle Cornish *gylwyr* 'is called', *gyllyr* 'one can', Middle Breton *mireur* 'is watched', *guilir* 'one sees' and Middle Welsh *byddir* 'one may be', *lladawr* 'will be killed'. When it comes to specific passive features, such as the use of 'by'-phrases, there is variation with Middle Breton and Welsh lacking by-phrases with impersonal morphology, where this may be found in Middle Cornish, as shown in (20). The use of the by-phrase in later Welsh and Breton appears to be a parallel innovation due to contact with English and French, respectively (Willis forthcoming).

(20) maythogh keris gans lues [Middle Cornish]
that love.IMP by many
'that you are loved by many' (BM 272)

Examples like the above in Cornish with *-is/-ys* can be ambiguous, however, between imperfect impersonal morphology and past participles with left-out auxiliary 'be'. Periphrastic passives with 'be' + past participle are also found in Breton,

which just like in Cornish is due to language contact (Lewis & Piette 1966: 49–50, Le Roux 1957: 125–126).

(21) a. prak gans pup na vethaf lethys [Middle Cornish]
why by all NEG be.1S slay.PPP
'why by all I may not be slain' (O.595-96)

b. ha pa ez vezo laqueat mat euez [Middle Breton]
and when PRT be.FUT.3S put.PPP good attention
'and when good attention has been (will be) paid...' (Cathell 5)

Middle Welsh, on the other hand, innovates a *cael*-passive, which is typologically similar to the English *get*-passive, although this was probably an internal innovation (Willis forthcoming). Baker & Roberts (forthcoming) call the verbal parameter that captures the distinction between periphrastic and non-periphrastic passives Pv38 (VSV) 'Voice spread to V', which essentially asks whether voice is realized both on V and separately. The subtle distinctions between different types of periphrastic constructions exhibited by British Celtic languages are not captured by any specific parameter yet, however. Note that opportunities for further extension of the parameters in the passive domain could be based on the suggestion that the *by*-phrase doubles the passive argument *-en* (Baker et al. 1989: 223), which is compatible with the observation that there is a distinction between transitive and unaccusative readings of, for example, 'cost' in Welsh when they are used in the impersonal form, as shown in the Modern Welsh examples in (22) (see Arman 2015: 133–136 and Roberts 2019: 425fn5 for discussion):

(22) a. costiwyd y CD yn ddeg-punt gan y cwmni.
cost.PAST.IMP the CD PRED ten-pounds by the company
'The CD was costed at 10 GBP by the company.'

b. *Costiwyd (yn) deg-punt (gan y CD).
cost.PAST.IMP PRED ten-pounds by the CD

All of the examples in this section so far do show that when we know enough about the history of languages and their historical contact situations, this can inform our syntactic reconstructions. British Celtic languages form an excellent case study for this since we have medieval data for all three and despite the fact that they are closely related, we do see clear syntactic differences as well. All of these differences can be due to transfer from French (in Breton) and English (in Cornish), because we know about the socio-historical situation in which these

languages came into contact. Furthermore, we know enough about the historical stages of both French and English to be able to answer questions regarding syntactic parameters like Pv38 on periphrastic passives, which is set positively in both Middle English and Middle French.

The PCM offers another possible way to explore cases of (potential) syntactic transfer. Since loanwords are essentially cast aside before applying the comparative method to reconstruct language families, the resulting tree topologies are supposed to show vertical inherited lineages only, neglecting data that could indicate language contact. This means that even in large-scale Bayesian applications of phylogenetic reconstruction based on large cognate databases, historical cases of language contact (or, 'horizontal transfer') can only be detected using additional measures (as done by, for example, Neureiter et al. 2022). Since there is no such filtering out of contact features in the first stage of syntactic reconstruction using the PCM, this means the resulting topologies of the reconstructed trees would include similarities due to transfer in their vertical branches if there were any. Therefore, if we compare the traditional phonological and syntactic tree topologies, we would then expect them to differ exactly in those cases where language contact played a role. Since syntactic transfer is often argued to be more difficult than lexical borrowing, we may not see as many differences, but still, family trees and in particular subfamilies would be predicted to look slightly different. Although other factors could play a role here too, this prediction is certainly borne out in both the nominal and the verbal sets of parameters, as shown by Baker & Roberts (forthcoming) in the case of Germanic in particular. Such instances of divergence between phonological and syntactic reconstructions force us to zoom into the exact cases of identity and difference in syntactic features, thus providing us with a detailed research program. In the next section I discuss more examples of how Minimalist syntactic reconstruction, and the PCM through the use of parameter hierarchies in particular, can help us gain more insight into (Celtic) language history.

4 How Minimalist syntactic reconstruction can help Celtic

Although we do have older attestations of some modern Celtic languages like Irish, Breton and Welsh, evidence is still scarce compared to other branches of the Indo-European language family. The first Old Irish and Old Welsh texts are of a much later date than the first attestations we have in Greek, Latin or Indo-Iranian, which is one of the reasons why Celtic has always lagged behind in providing evidence for Proto-Indo-European reconstruction. In addition, various

(stages of) Celtic languages are not even attested at all, such as the language spoken by the first Celtic speakers moving to the British Isles or any of the modern languages that could have developed from later stages of Gaulish, Celtiberian or Lepontic. Finally, extant evidence for any of the Continental Celtic languages is scarce, and mainly consists of short inscriptions, which severely restricts the possibilities of getting a comprehensive picture of the overall grammatical system of any of these languages. Issues like these are further complicated by ambiguous and sometimes contradictory evidence from historical phonology. Some scholars (e.g. Kortlandt 2007; Schrijver 1997 and more recently Weiss 2022) have postulated an Italo-Celtic branch of Indo-European, based on certain similarities in phonology and morphology such as sigmatic future with reduplication, substitution of **-bhos* for PIE dat.PL **-mus* and the lack of innovation of an elaborate middle voice. More recent Bayesian approaches based on lexical cognates often place Celtic and Romance languages close together as well, e.g., Atkinson & Gray (2005: 215) and Neureiter et al. (2022: 7). However, even Kortlandt (2007) who presents a concrete reconstruction of Proto-Italo-Celtic, acknowledges that "specific Italo-Celtic innovations are few" (Kortlandt 2007: 157), which has led others (e.g., Watkins 1966) to reject this Italo-Celtic hypothesis at all originally (see Weiss forthcoming, however, for shared morphosyntactic features of preverbs and adpositions). Similarly, intermediate subgroups within Celtic itself are not always easy to establish. Insular Celtic languages like Welsh and Irish exhibit a number of similarities (Schrijver 1995: 463–465), but in some respects the British Celtic languages appear much closer to Gaulish (e.g., as a P-Celtic language) leading scholars like Sims-Williams (2003) to hypothesize that "British" could be inherited from both Gallo-Brittonic as well as Insular Celtic.

In both of these cases comprehensive methods for syntactic reconstruction like the PCM offer opportunities to fill the gaps in our linguistic history. This is not just apparent for phylogenetic topology on a larger scale, but also for specific gaps in our knowledge of syntactic features of Celtic languages on a smaller scale. In this final section, I first focus on syntactic changes in the history of some insular Celtic languages specifically, to create a better picture of the grammars of intermediate stages like Proto-British, but also to identify gaps in our knowledge and opportunities for future research. Finally, I discuss examples of syntactic features shared by Celtic and Romance, as well as where they differ based on a comparison of modern languages like present-day Irish and Welsh, on the one hand, and French, Spanish and both Northern and Southern Italian varieties, on the other, which leads to a final discussion of Celtic as a subfamily of Proto-Indo-European in general.

4.1 Reconstructing intermediate stages

In Section 3.2, we saw examples of syntactic innovations in the Voice system of British Celtic languages: both Middle Cornish and Breton already show examples of the *be*-passive, whereas Late Middle Welsh innovated the *cael*-passive, which behaves more like the English *get*-passive. Old and Middle Irish, on the other hand, did not have periphrastic passives, which would leave the verbal parameter Pv38 (VSV) in the negative setting as there is no evidence for a Voice feature being realized on both the verb and separately with an auxiliary. Present-day Irish may have innovated a periphrastic passive too, but this is so far limited to the perfective passive only (McCloskey 1996: 254–255). Since the periphrastic passives are a later innovation in Cornish, Breton and especially Welsh (all of which still have morphological passives and impersonals), we should reconstruct the negative setting of this parameter for Proto-British as well as Proto-Insular-Celtic (if this is an intermediate stage we want to reconstruct at all). The potential positive setting for present-day Irish depends on the exact interpretation of the verb 'be' as a possible auxiliary, which is debatable and it may mean this parameter is still set to negative even today. Note that this discussion reveals two important points about the PCM. First, the restriction to perfective passives in present-day Irish could lead us to refine the passive parameter hierarchy to allow for this type of microvariation. Second, it was the parametric questions and the comparison of the cross-linguistic settings, which highlighted this gap in our knowledge about Irish to begin with.[5]

A further example of systematic comparison driven by the PCM can be found in the domain of negation. Baker & Roberts (forthcoming) posit Pv83 (NMU) as the parameter for "Multiple Negation", which essentially asks whether a language permits negative concord. Modern Irish and Welsh differ in this respect, as shown in (23) and (24), with only the latter overtly realizing this feature thus rendering a positive setting.

(23) Níor chreideas é [Modern Irish]
NEG believe.PAST.1S 3SM.ACC
'I did not believe him.' (Nolan 2018)

(24) a. D-oes neb yn ennill. [Modern Welsh]
NEG-be.3S no.one PROG win.VN

[5]Many thanks to Jim McCloskey and Ian Roberts for sharing their discussion notes, which brought this to my attention.

b. *Mae neb yn ennill.
be.3s no.one prog win.VN
'No one is winning.' (Willis 2013)

Middle Welsh did not have negative concord, on the other hand, predominantly showing Stage I of Jespersen's Cycle with preverbal negation only, shown in (25a). Middle Breton already appears to be in Stage II showing both preverbal negation as well as postverbal *ket*, shown in (25b), which has been spreading within the language as a Breton-specific innovation probably due to French (Hemon 1975: 281–286; Poppe 1995: 103–105; Willis 2006, forthcoming).

(25) a. **ny** wnn i pwy wyt-ti [Middle Welsh]
NEG know.pres.1s I who be.pres.2s-you
'I don't know who you are.' (PKM 2.22-3)

b. Ne chome **ket** pell e wenneien en e c'hodell [Middle Breton]
NEG stay.3s **ket** long 3SM pennies in 3SM pocket
'His pennies did not stay long in his pockets.' (MAV p.29)

Cornish, just like Irish in this case, only allows a preverbal marker *ny* or *neng/nyngs* (before forms of *bos* 'be' and *mones* 'go'), with or without preverbal subject, as shown in (26).

(26) a. **Ny** lowenhaf in ow dythyow neffra lam [Cornish]
NEG rejoice.1s in 1s days never now
'I shall not rejoice ever now in my days.' (BK 2929–2930)

b. **Nyng** ew ow thowl
NEG be.3s 1s intention
'It is not my intention.' (BK 293)

c. My **ny** won. / My a wor
I NEG know.1s / I PTC know.3s
'I don't know.' (PA 121a / OM 185)

These examples reveal another curious fact about Cornish, namely that verbs can exhibit agreement with pronominal subjects that precede the negation (Lewis 1956: 48–49(§47)), which is impossible in their positive counterparts. Pronouns cannot be overtly expressed in Cornish, or most other Celtic languages, if there is subject-verb agreement (i.e., the so-called "complementarity principle", Stump 1984: 292, Borsley & Stephens 1989; Doron 1988). However, this principle clearly

does not hold for subject-initial negative clauses in Cornish. It also does not hold for Welsh, although in post-verbal position, only the weaker forms of the pronouns (echo pronouns) can be used, as shown in (27a). In Middle Welsh, strong pronouns can be found with agreement, but crucially only when they occur in preverbal position, as shown in (27b). When it comes to nominal subject DPs, they can never agree with verbs in number (shown in 27c), which Roberts (2019: 388–389) relates to the 3rd-person forms specifically.

(27) a. dywedais i/*fi [Modern Welsh]
say.PAST.1S I.WEAK/STRONG
'I said'

b. mi a wnn pwy wyt ti ac ny chyuarchaf i [Middle Welsh]
I.STRONG PTC know.1S who are.2S you and NEG greet.1S I.WEAK
'I know who you are, but I will not greet you.' (Pwyll p1rc2l18)

c. *Canon / Canodd y plant. [Modern Welsh]
sing.PAST.3PL / sing.PAST.3S the children.PL
'The children sang.'

These subtle distinctions among the Celtic languages in terms of agreement and the link to negation in Cornish in particular reveal gaps in the grammars of the intermediate proto-languages we would like to reconstruct, showing us exactly where further research is necessary. When it comes to the use of multiple negation, however, it is clear that negative concord was an innovation in Welsh and Breton only. This cannot be reconstructed to Proto-British as they each exhibit different postverbal negative markers (Welsh *ddim* vs. Breton *ket*), which are not lexical cognates. They do share the same syntactic features, however, but Willis (2006) has shown that the changes in these features can be traced exactly through the history of Welsh. For Proto-British, as well as Proto-Insular-Celtic, we therefore reconstruct the negative value for the Multiple Negation parameter, which is already found in Cornish, Irish, as well as Old and Middle Welsh. Note that this is also in line with the negative setting of this parameter in other Indo-European languages like Latin, Vedic and Hittite, which brings us to comparing Celtic parameters beyond the British Isles. Since evidence for Continental Celtic languages is scarce, the task of reconstructing its syntax seems daunting, especially from the perspective of reconstruction methods that rely on overt (lexical) cognacy of functional items as there are very few (if any) such items that can be found in both Continental and Insular Celtic languages. The PCM offers an

opportunity here too since it not only provides a systematic approach seeking concrete answers to a fixed set of parametric questions in the nominal and clausal domains, it also allows for reconstruction based on more covert realizations, such as movement or Internal Merge due to requirements of specific features on functional heads. In addition, the strong hierarchical structure with built-in implicational parameters means that a number of parameter values can be inferred from the positive or negative setting of others. Combining these strengths of the PCM will provide experts on Continental Celtic with the opportunity to reconstruct a grammar of the proto-language that will cover more aspects of syntax than was ever possible before.

4.2 Syntactic similarities in Italo-Celtic

In this final section, I show examples of syntactic features that can be reconstructed for Celtic, Romance and/or both to shed light on the "Italo-Celtic" hypothesis from a new angle of Minimalist syntax. Obvious similarities are found for fundamental (macro)parameters, such as using bound morphemes (Pn1 FMG "grammaticalized morphology"), agreement (Pn2 FGA "grammaticalized agreement") and Case (Pn3 FGK "grammaticalized Case"), which tend to distinguish Indo-European languages from languages like Mandarin, Japanese and Wolof. Much more interesting for our purposes are the similarities and differences found within the Indo-European family and languages these came into contact with in the course of history, such as Basque, Hungarian and a range of Semitic languages.

When focusing on Celtic, Romance and Germanic in particular, there are various parameters with positive settings shared with some (or all) languages in those three subfamilies, such as grammaticalized person, number and gender in both the nominal and verbal domains. In addition, all three subfamilies share features to do with the Categorization Requirement (CR, cf. Panagiotidis 2015), which states that every c-selecting and c-selected functional head H must be uniquely identified in relation to an EP (Extended Projection). It is built on the assumption that non-categorizing heads do not have intrinsic V- or N-features, but instead bear FF [Cat:__]. The higher heads in the EP then are categorially identified as belonging to that EP. Roberts (2019: 164–166) argues that UG makes available three mechanisms to satisfy the Categorization Requirement, yielding three types with Type I satisfying CR through roll-up, labeling all the functional categories as V or N giving rise to harmonic head-finality. Type II, on the other hand, satisfies CR through head-movement yielding head-initial languages, and, Type III satisfies CR by lexicalizing heads in the EP. The "purest" versions of these three types are rare and there is a possibility to combine types for different parts

of the syntax. Celtic, Romance as well as English combine Types II and III, but other West Germanic languages can also feature Type I, e.g., German featuring all three types: roll up of VP around v, head movement as well as lexicalizing heads. Roberts (2019: 163) argues that both Welsh and English allow intrinsic labeling followed by head-movement, as shown, for example, by the inflected auxiliary *oedd* in example (28) raising to a high functional position (e.g., some C head) when there are free lower aspectual heads, which could be considered some sort of "*be*-insertion" (cf. English dummy "*do*-support"):

(28) Oedd y bachgen wedi bod yn ymlad [Welsh]
be.3S.PAST the boy PERF be.INF PROG fight.INF
'The boy had been fighting.'

Syntactic features that are even more interesting for investigating Celtic as a subfamily are those that distinguish the order of adjectives and nouns, for example, NM1 "N under M1 Adjectives". Given the cross-linguistically canonical sequence of structured adjectives, Ceolin et al. (2021: 41) say that "Manner1 adjectives" (i.e., those describing quality or size as opposed to shape or color, which are 'Manner2') can surface to the left of the noun in languages like Italian, French, Spanish, Germanic, Slavic language family and Standard Greek. In Welsh, Irish, Farsi or some Romance dialects of Italy, however, they cannot, as shown in (29):

(29) a. un bel nuovo vestito blu [Italian]
a nice new dress blue

b. ffrog las newydd neis [Modern Welsh]
dress blue new nice
'a nice new blue dress'

Although these word order facts are relatively clear, agreement with adjectives in Celtic is not as straightforward as it seems, which becomes apparent when asking these detailed parametric questions, especially for older stages of the Celtic languages. Borsley et al. (2007) claim that "[I]n predicative position, adjectives never agree in gender or number with their noun" (2007: 179), citing example (30):

(30) Mae ei lygaid yn las / *leision. [Modern Welsh]
be.3S 3SM eyes PRED blue.GENERAL / blue.PL
'His eyes are blue.'

In spoken Welsh corpora, however, it is possible to find examples, as shown in (31):

(31) roeddynt yn fychain [Modern Welsh]
were.3PL PRED small.PL
'they were small' (CEG corpus; Meelen & Nurmio 2020)

Meelen & Nurmio (2020) show that agreement in earlier stages of Welsh was different, as shown in (32), which is backed up by a corpus study of Middle and Early Modern Welsh with 17 out of 50 predicative adjectives (34%) showing agreement. Since 13 out of the 17 examples stem from the Bible translation, some further research is necessary here before we can definitively show a parametric change, however.

(32) Mae y cyflenwadau o wenith Lloegr a Thramor yn fychain
be.3S the supplies of wheat English and overseas PRED small.PL
'the supplies of English and overseas wheat are small' (Papurau Newydd: Yr Amserau 12 Nov 1851)

Other examples of parameters that put Welsh and Irish close to the majority of Romance, but also close to Germanic languages are Grammaticalized Aspect (GRA) and Mood (GRM) and Mood- and Voice-checking (MCV and VCV) in the verbal domain as well as Grammaticalized Specified Quantity (DGR) in the nominal domain (encoding definiteness).

There are, however, also various parameters that show differences between Celtic and Romance languages, especially when we zoom in to the micro- and nano-levels. Roberts (2019) argues that verb-initial languages like Modern Welsh and Irish copy φ-features (i.e., Person/Gender/Number features) of the verbal suffix in one of the other agreeing heads (enclisis), just like Northern Italian Dialects (NIDs), where this is shown by subject (pro)clitics. Unlike NIDs, the Welsh weak/clitic pronouns actually combine both Person and Number features, creating a microparametric distinction. Other Romance varieties differ even more from Celtic in this respect, since they do not exhibit agreeing subject clitics (either enclitic or proclitic) altogether. Similarly, Welsh and Irish differ from languages like French and Brazilian Portuguese by optionally allowing the suppression of φ-features on T, making them into Partial Null-Subject languages (Roberts 2019: 608). Other Romance languages go one step further, not having obligatory EPP features on T at all, yielding them as Canonical Null-Subject languages with fully specified Person-features found on both T and D heads.

In the nominal domain, Celtic languages pattern with Semitic and Icelandic freely admitting bare singular count indefinite arguments (nominal Pn19 "Weak Specified Quantity", CGR), e.g., Hebrew *kelev*, Welsh *ci* '**a** dog', where Romance

and many Germanic languages require an indefinite article, e.g. Dutch *een hond*, French *un chien* and Italian *un cane* 'a dog'. Similarly parameter Pn47 (GFL) groups Celtic languages together with Greek, most Slavic languages as well as German and Icelandic in which a non-adpositional, non-iterable Genitive Case appears to the right of canonically ordered adjectives, as shown in (33) from Ceolin et al. (2021: 31):

(33) portread hardd y plentyn [Modern Welsh]
portrait beautiful the child
'the child's beautiful portrait'

Finally, there are also parameters that set Welsh and Irish apart from not just Romance, but most of Indo-European, such as Pn55 "plural spread from cardinal quantifiers" (PSC). In Welsh, for example, singular nouns are used after cardinal quantifiers, shown in (34), rendering a negative setting for this parameter, similar to Farsi, Uralic and Turkic languages (but unlike English or the rest of Indo-European).

(34) un ffordd / tri ffordd [Modern Welsh]
one way / three way
'one way / three ways'

If we combine the results from 87 verbal (Baker & Roberts forthcoming) and 94 nominal parameters (Ceolin et al. 2021), the first observation is that languages that are usually grouped together into subfamilies using the comparative method of phonological reconstruction also form subfamilies based on syntactic features. Within Indo-European, we have a Slavic branch with Russian, Polish, Serbo-Croatian and Slovenian; a Germanic branch with Norwegian, Danish, German, Dutch, English, Afrikaans and Icelandic; a Romance branch with Spanish, Portuguese, Italian and French; and Modern Irish and Modern Welsh are grouped together representing Celtic. The Hamming distance between those two modern Celtic languages, which counts both positive and negative settings that are similar as "identities", is 0.039; the Jaccard distance, which ignores identical settings if they are both negative, is 0.167.[6] In general, the phylogenetic tree topologies

[6]Note that Baker & Roberts (forthcoming) only provide Hamming distances, whereas Ceolin et al. (2021) change their previous Hamming metric to Jaccard to be able to distinguish negative similarities from positive ones. Ceolin et al. (2021) report no difference at all, i.e., $\delta = 0.0$ between Irish and Welsh. Although the Insular Celtic DP indeed shows striking similarities, it is worth noting that 100% similarity in 94 distinct parameters is very unusual even for languages that are very closely related. In fact it is the only language pair in their dataset that shows 0 differences for relevant comparanda.

based on Hamming and Jaccard distances of all parameters combined are quite similar, although the Jaccard distances are about 0.13–0.2 larger. Note that with either metric, Irish and Welsh clearly are below the $\delta < 0.20$ threshold Bortolussi et al. (2011) suggested for probable relatedness.

If we compare both Irish and Welsh combined to other subfamilies, we see the average distance distributions in Table 1. Celtic, as represented by Irish and Welsh, on average shares more syntactic features with the Slavic languages than with Romance: both average Hamming and Jaccard distances are lower for Slavic than for Romance. If the parameters would solely be used for the purposes of detecting historical relatedness, this is a somewhat surprising result since family trees based on the traditional phonological comparative method would often posit a combined Italo-Celtic subgroup (Weiss 2022). Weiss (forthcoming), for example, shows that shared innovations and retentions in the system of preverbs and adpositions provide additional evidence for this closer relation between the Celtic and Romance languages (see also Section 4.2).

Table 1: Hamming and Jaccard distances for averaged subfamilies

	Hamming	Jaccard
Welsh vs Romance	0.204	0.364
Welsh vs Germanic	0.218	0.395
Irish vs Romance	0.196	0.358
Irish vs Germanic	0.227	0.429
Celtic vs Romance	0.200	0.361
Celtic vs Germanic	0.227	0.412
Celtic vs Slavic	0.185	0.323
Celtic vs Uralic	0.262	0.468
Celtic vs Semitic	0.233	0.400

This highlights an important point of interpretation and application of using the PCM for a variety of purposes. First, the results of detecting phylogenetic signals and establishing historical relations between subgroups should not necessarily be expected to be the same not only because lexical and phonological features on the one hand and grammatical features on the other are representing different parts of the language, but also because potential evidence for horizontal transfer (i.e., transfer due to contact, see Section 3.2) is systematically excluded for lexical/phonological reconstruction, but not for the grammatical features in

the PCM. Second, the number and range of languages and features representing a subgroup is important in any statistical reconstruction method (whether Bayesian or not). The Celtic subgroup is only represented by two languages in the PCM and as they are less-studied some features in the nominal and verbal domain still have some question marks. The Romance languages, on the other hand, are all studied in a high amount of detail and are represented even in its smallest iteration by at least five languages (French, Italian, Spanish, Portuguese and Romanian). Slavic languages currently fall somewhere in between, with four languages represented (Slovenian, Serbo-Croatian, Russian and Polish), some of which are well-studied from a syntactic point of view. Finally, it is worth bearing in mind that only the nominal and verbal parameters are part of the dataset yielding the above results. When clausal parameters that cover syntactic features of the C-domain are included, the overall results may look very different. Since Longobardi & Guardiano (2009) note that the nominal parameters are more likely to be robust over time, it is worthwhile zooming in on potential changes within the Celtic subfamily in the verbal domain. Modern Irish and Welsh show a Jaccard $\delta = 0.263$, which is much higher than the average of both nominal and verbal parameters since Ceolin et al. (2021) report 0 difference in the nominal domain. If we compare Middle Welsh to Old Irish, on the other hand, there are fewer syntactic distinctions with Jaccard $\delta = 0.171$. Between Middle and Modern Welsh verbal parameters Jaccard $\delta = 0.268$, whereas Irish appears to have changed fewer verbal parameters showing Jaccard $\delta = 0.229$.

Overall, these numbers are preliminary. As the number of parameters increases, we can expect to see a refinement of the analysis. Despite this, it is clear that subfamilies can be easily identified already and we can identify gaps in our knowledge of the old and modern Celtic languages. Even if the PCM eventually were to yield different results related to phylogenetic signal and subgrouping, we can see that it is a fruitful exercise for reconstructing the grammatical system of proto-languages.

When comparing parameter settings in both the modern and the medieval Celtic languages, we can get a much better picture of the syntactic developments going back to Proto-British and Proto-Celtic. For every reconstructed parameter whose values are different from those in the modern daughter languages, we should carefully address some of the remaining issues of syntactic reconstruction discussed in Section 2, e.g., the radical-reanalysis and transfer problems. The latter was already discussed in Section 3.2 and requires knowledge about the historical contact situation, as well as the application of methods to detect syntactic transfer in particular, such as those described by Bowern (2008) and Daniels (2017). The radical-reanalysis problem can only be addressed applying

the careful step-by-step methodology presented by Willis (2006), which involves a critical analysis of how every stage of reanalyses and extensions could have led to the syntactic patterns we find in attested stages of all the daughter languages. This is more difficult when we are forced to reconstruct negative values of parameters, since they do not have overt realizations, forcing us to rely more on implicational generalizations known from (Minimalist) theoretical syntax such as those worked out in detailed parameter hierarchies. A large number of parameters, however, (at least around a third for the verbal domain[7]) actually ask about direct PF realizations such as inflectional morphology, which can potentially be reconstructed using the traditional comparative method if they are true cognates, as done by Walkden (2011, 2014) and Willis (2011).

5 Conclusion

In this chapter, I have shown that recent advances in Minimalist syntactic theory enable us to tackle the long-standing challenges of syntactic reconstruction. While various methods of syntactic reconstruction have been put forward in the last decades, Minimalist approaches that rely on the Borer-Chomsky Conjecture (BCC) have been particularly successful in addressing problems of directionality, transfer and radical reanalysis. The Parametric Comparison Method (PCM) has some additional benefits of providing a comprehensive overview of the syntax of a proto-language because of the discrete nature of parameters, at the same time avoiding the need to find potentially improbably phenomena (e.g., sound correspondences in restricted contexts) and reliance on vague cognacy judgments (e.g., due to semantic changes). The PCM is most powerful when it is combined with the careful step-by-step evaluation of the reanalyses and extensions in all reconstructed stages shown by Willis (2011) and methods for detecting language contact in the domain of syntax (Bowern 2008, Daniels 2017).

As I showed in Section 3.1, there are various ways in which the parametric questions can be improved and extended, but the language family trees based on nominal or verbal domains (or both) show remarkable resemblances to the tree topologies based on lexical cognates using the traditional comparative method. Celtic languages can play an important role in the further development of the PCM due to their unique properties in syntactic features, even though only Modern Welsh and Irish have been taken into account so far. This knowledge can help us refine and extend the PCM, on the one hand, but on the other, the PCM can help our understanding of historical syntax by providing a systematic set of questions that can shape a clearly-defined research program for the Celtic languages.

[7] Jim Baker, p.c.

References

Alexopoulou, Theodora. 2006. Resumption in relative clauses. *Natural Language and Linguistic Theory* 24(1). 57–111. DOI: 10.1007/s11049-005-0898-2.

Arman, Laura. 2015. *The Welsh impersonal construction*. University of Manchester. (Doctoral dissertation).

Atkinson, Quentin D. & Russell D. Gray. 2005. Curious parallels and curious connections—phylogenetic thinking in biology and historical linguistics. *Systematic Biology* 54. 513–26. DOI: 10.1080/10635150590950317.

Auroux, Sylvain. 1990. Representation and the place of linguistic change before comparative grammar. In Tullio De Mauro & Lia Formigari (eds.), *Leibniz, Humboldt, and the origins of comparativism*, 213–238. Amsterdam: John Bejamins. DOI: 10.1075/sihols.49.15aur.

Baker, Jim & Ian Roberts. Forthcoming. Extending parametric comparison: Preliminary results. *Journal of Historical Syntax*. DOI: 10.33774/coe-2021-6lzqq.

Baker, Mark. 1996. *The polysynthesis parameter*. Oxford: Oxford University Press.

Baker, Mark. 2008. The macroparameter in a microparametric world. In Theresa Biberauer (ed.), *The limits of syntactic variation*, vol. 132, 351–373. Amsterdam: John Benjamins.

Baker, Mark, Kyle Johnson & Ian Roberts. 1989. Passive arguments raised. *Linguistic Inquiry* 20(2). 219–251.

Balles, Irene. 2008. Principles of syntactic reconstruction and "morphology as paleosyntax". In Gisella Ferraresi & Maria Goldbach (eds.), *Principles of syntactic reconstruction*, 161–186. Amsterdam: John Benjamins. DOI: 10.1075/cilt.302.08bal.

Barðdal, Jóhanna. 2013. Syntax and syntactic reconstruction. In *The Routledge handbook of historical linguistics*, 343–373. Routledge.

Barðdal, Jóhanna & Thórhallur Eythórsson. 2012. "Hungering and lusting for women and fleshly delicacies": Reconstructing grammatical relations for Proto-Germanic. *Transactions of the Philological Society* 110(3). 363–393. DOI: 10.1111/j.1467-968x.2012.01318.x.

Barðdal, Jóhanna & Spike Gildea. 2015. Diachronic construction grammar: Epistemological context, basic assumptions and historical implications. In Jóhanna Barðdal, Elena Smirnova, Lotte Sommerer & Spike Gildea (eds.), *Diachronic construction grammar*, 1–50. Amsterdam: John Benjamins.

Barðdal, Jóhanna, Spike Gildea & Eugenio Lujan. 2020. *Reconstructing syntax*. Leiden: Brill.

Beekes, Robert S. P. 1995. *Comparative Indo-European linguistics: An introduction*. Amsterdam: John Benjamins.

Berwick, Robert C. & Noam Chomsky. 2011. The biolinguistic program: The current state of its development. In Anna-Maria di Sciullo & Cedric Boeckx (eds.), *The biolinguistic enterprise*, 19–41. Oxford: Oxford University Press.

Biberauer, Theresa. 2019. Factors 2 and 3: Towards a principled approach. *Catalan Journal of Linguistics*. 45–88. DOI: 10.5565/rev/catjl.219.

Biberauer, Theresa, Anders Holmberg, Ian Roberts & Michelle Sheehan. 2014. Complexity in comparative syntax: The view from modern parametric theory. In Frederick J Newmeyer & Laurel B Preston (eds.), *Measuring grammatical complexity*, 103–127. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199685301.003.0006.

Biberauer, Theresa & Ian Roberts. 2012. *The significance of what hasn't happened.* Lisbon: Presentation at Diachronic Generative Syntax 14 conference, Lisbon.

Biberauer, Theresa & Ian Roberts. 2017. Parameter setting. In *Cambridge handbook of historical syntax*, 134–62. Cambridge: Cambridge University Press.

Biberauer, Theresa & George Walkden (eds.). 2015. *Syntax over time: Lexical, morphological, and information-structural interactions.* Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199687923.001.0001.

Boeckx, Cedric. 2015. Some thoughts on biolinguistics. *Veritas (Porto Alegre)* 60(2). 207–221. DOI: 10.15448/1984-6746.2015.2.22006.

Borsley, Robert D. & Janig Stephens. 1989. Agreement and the position of subjects in Breton. *Natural Language & Linguistic Theory* 7(3). 407–427.

Borsley, Robert D., Maggie Tallerman & David Willis. 2007. *The syntax of Welsh.* Cambridge: Cambridge University Press.

Bortolussi, Luca, Andrea Sgarro, Giuseppe Longobardi & Cristina Guardiano. 2011. How many possible languages are there? *Biology, computation and linguistics* 228. 168–179. DOI: 10.3233/978-1-60750-762-8-168.

Bošković, Željko. 2008. What will you have, DP or NP? In *Proceedings ofNELS*, vol. 37:1, 101–14. New Jersey.

Bouckaert, Remco, Philippe Lemey, Michael Dunn, Simon J Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard & Quentin D. Atkinson. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337(6097). 957–960. DOI: 10.1126/science.1219669.

Bowern, Claire. 2008. Syntactic change and syntactic borrowing in generative grammar. In Gisella Ferraresi & Maria Goldbach (eds.), *Principles of syntactic reconstruction*, 187–216. Amsterdam: John Benjamins. DOI: 10.1075/cilt.302.09bow.

Brandi, Luciana & Patrizia Cordin. 1989. Two Italian dialects and the null subject parameter. In Oxvaldo Jaeggli & Kenneth J. Safir (eds.), *The Null Subject Parameter*, 111–42. Dordrecht: Kluwer.

Campbell, Lyle. 2013. *Historical linguistics: An introduction*. 3rd edn. Edinburgh: Edinburgh University Press.

Campbell, Lyle & Alice C. Harris. 2002. Syntactic reconstruction and demythologizing "Myths and the prehistory of grammars". *Journal of Linguistics* 38. 599–618. DOI: 10.1017/s0022226702001706.

Carnie, Andrew, Heidi Harley & Elizabeth Pyatt. 2000. VSO order as raising out of ip? In Andrew Carnie & Eithne Guilfoyle (eds.), *The syntax of verb initial languages*. Oxford: Oxford University Press. DOI: 10.1093/oso/9780195132229.003.0003.

Ceolin, Andrea, Cristina Guardiano, Monica Alexandrina Irimia & Giuseppe Longobardi. 2020. Formal syntax and deep history. *Frontiers in Psychology* 11(488871). DOI: 10.3389/fpsyg.2020.488871.

Ceolin, Andrea, Cristina Guardiano, Giuseppe Longobardi, Monica Alexandrina Irimia, Luca Bortolussi & Andrea Sgarro. 2021. At the boundaries of syntactic prehistory. *Philosophical Transactions of the Royal Society B: Biological Sciences* 376(1824). rstb.2020.0197, 20200197. DOI: 10.1098/rstb.2020.0197.

Chomsky, Noam. 1980. On binding. *Linguistic Inquiry* 11(1). 1–46.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge: MIT Press.

Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36(1). 1–22. DOI: 10.1162/0024389052993655.

Clackson, James. 2017. The comparative method and comparative reconstruction. In Adam Ledgeway & Ian Roberts (eds.), *The Cambridge handbook of historical syntax*, 189–206. Cambridge: Cambridge University Press. DOI: 10.1017/9781107279070.010.

Clark, Robin & Ian Roberts. 1993. A computational model of language learnability and change. *Linguistic Inquiry* 24(2). 299–345. DOI: 10.4324/9781315310572.

Crisma, Paola, Cristina Guardiano & Giuseppe Longobardi. 2020. Syntactic diversity and language learnability. *Studi e Saggi Linguistici* 58(2). DOI: 10.4454/ssl.v58i2.265.

Croft, William. 2001. *Explaining language change: An evolutionary approach*. Oxford: Oxford University Press.

Currie, Oliver. 2013. Gradual change and continual variation: The history of a verb-initial construction in Welsh. In Anna Giacalone Ramat, Caterina Mauri & Piera Molinelli (eds.), *Studies in language companion series*, vol. 133, 43–78. Amsterdam: John Benjamins. DOI: 10.1075/slcs.133.03cur.

Daniels, Don. 2017. A method for mitigating the problem of borrowing in syntactic reconstruction. *Studies in Language* 41(3). 577–614. DOI: 10.1075/sl.41.3.02dan.

Delbrück, Berthold. 1893. *Vergleichende Syntax der indogermanischen Sprachen.* Strasbourg: E. K. J. Trübner.

Doron, Edit. 1988. On the complementarity of subject and subject-verb agreement. In Michael Barlow & Charles Ferguson (eds.), *Agreement in natural language: Approaches, theories, descriptions*, 201–218. Stanford: CSLI Stanford.

Douglas, Jamie. 2017. Unifying the that-trace and anti-that-trace effects. *Glossa* 2(1):60). 1–28. DOI: 10.5334/gjgl.312.

Dressler, Wolfgang. 1969. Eine textsyntaktische Regel der idg. Wortstellung: (Zur Anfangsstellung des Prädikatverbums). *Zeitschrift für vergleichende Sprachforschung* 83(1). 1–25.

Dryer, Matthew. 2007. Review of Frederick J. Newmeyer. possible and probable languages: A generative perspective on linguistic typology. *Journal of Linguistics* 43. 244–52. DOI: 10.1017/s0022226706254586.

Dunn, Michael, Angela Terrill, Ger Reesink, Robert A. Foley & Stephen C. Levinson. 2005. Structural phylogenetics and the reconstruction of ancient language history. *Science* 309(5743). 2072–2075. DOI: 10.1126/science.1114615.

Durie, Ross & Malcolm Ross. 1996. *The Comparative method reviewed: Regularity and irregularity in language change.* Oxford: Oxford University Press.

Dyen, Isadore, Joseph Kruskal & Paul Black. 1992. An Indo-European classification: A lexicostatistical experiment. *Transactions of the Philological Society* 82. 1–132.

Eska, Joseph F. 1996. On syntax and phonology within the Early Irish verbal complex. *Diachronica* 13. 225–257. DOI: 10.1075/dia.13.2.03esk.

Eska, Joseph F. & Benjamin Bruch. 2021. Remarks on pragmatic fronting and poetic overdetermination in Middle Cornish. *North American Journal of Celtic Studies* 5. 131–193. DOI: 10.1353/cel.2021.0014.

Eska, Joseph F. & D. Ellis Evans. 1993. Continental Celtic. In Martin Ball & Nicole Müller (eds.), *The Celtic languages*, 26–63. London: Routledge.

Eythórsson, Thórhallur & Jóhanna Barðdal. 2016. Syntactic Reconstruction in Indo-European: State of the Art. *Veleia* 33. 83–102. DOI: 10.1387/veleia.16823.

Ferraresi, Gisella & Maria Goldbach. 2008. *Principles of syntactic reconstruction.* Amsterdam: John Benjamins. DOI: 10.1075/cilt.302.

Fillmore, Charles J., Paul Kay & Mary Catherine O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language* 64. 501–538. DOI: 10.2307/414531.

Fox, Anthony. 1995. *Linguistic reconstruction: an introduction to theory and method*. Oxford: Oxford University Press.
Friedrich, Paul. 1975. *Proto-Indo-European Syntax*. Washington: Institute for the Study of Man.
Gildea, Spike. 2000a. On the genesis of the verb phrase in Cariban languages. In Spike Gildea (ed.), *Reconstructing grammar: Comparative linguistics and grammaticalization*, 65–106. Amsterdam: John Benjamins.
Gildea, Spike (ed.). 2000b. *Reconstructing grammar*. Amsterdam: John Benjamins.
Gildea, Spike. 2002. Reconstructing Pre-Proto-Tupi-Guarani main clause grammar. In Ana Suelly Cabral & Aryon Rodrigues (eds.), *Atas do encontro internacional de pesquisadores de línguas indígenas*, 315–326. Universidade Federal do Para Belem.
Gildea, Spike, Eugenio R. Luján & Jóhanna Barðdal. 2020. The curious case of reconstruction in syntax. In *Reconstructing syntax*, 1–44. Leiden: Brill.
Givón, Talmy. 1971. Historical syntax and synchronic morphology: An archaeologist's field trip. *Chicago Linguistic Society* 7. 394–415.
Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
Graves, Eugene Van Tassel. 1962. *The Old Cornish Vocabulary*. Ann Arbor: Columbia University.
Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. *Universals of language* 2. 73–113.
Guardiano, Cristina & Giuseppe Longobardi. 2005. Parametric comparison and language taxonomy. In *Grammaticalization and parametric variation*, 149–174. Oxford: Oxford University Press.
Hammarström, Harald, Philipp Rönchen, Erik Elgh & Tilo Wiklund. 2019. On computational historical linguistics in the 21st century. *Theoretical Linguistics* 45(3-4). 233–245. DOI: 10.1515/tl-2019-0015.
Harlow, Stephen. 1981. Government and relativization in Celtic. In Frank Heny (ed.), *Binding and filtering*, 213–254. MIT Press.
Harris, Alice C. 2008. Reconstruction in syntax: Reconstruction of patterns. In Gisella Ferraresi & Maria Goldbach (eds.), *Principles of syntactic reconstruction*, 73–95. Amsterdam: John Benjamins. DOI: 10.1075/cilt.302.05har.
Harris, Alice C. & Lyle Campbell. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.
Harrison, Shelly. 2003. On the limits of the Comparative Method. In Brian D. Joseph & Richard D. Janda (eds.), *Handbook of historical linguistics*, 213–243. Oxford: Blackwell.

Hartmann, Frederik & George Walkden. 2024. The strength of the phylogenetic signal in syntactic data. *Glossa* 9(1). DOI: 10.16995/glossa.10598.

Hemon, Roparz. 1975. *A historical morphology and syntax of Breton*, vol. 3. Dublin: Dublin Institute for Advanced Studies.

Hendrick, Randall. 2000. Celtic initials. In Andrew Carnie & Eithne Guilfoyle (eds.), *The syntax of verb initial languages*, 13–37. Oxford: Oxford University Press. DOI: 10.1093/oso/9780195132229.003.0002.

Hock, Hans Henrich. 1991. *Principles of historical linguistics*. 2nd edn. New York: Mouton De Gruyter.

Holden, Claire. 2002. Bantu language trees reflect the spread of farming across sub-Saharan Africa: A maximum-parsimony analysis. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 269(1493). 793–799. DOI: 10.1098/rspb.2002.1955.

Huang, C. T. James. 2015. On syntactic analyticity and parametric theory. In Y.-H. Audrey Li, Andrew Simpson & W.-T. Dylan Tsai (eds.), *Chinese syntax in a cross-linguistic perspective*, 1–47. Oxford: Oxford University Press.

Humphreys, Humphrey L. 1992. The Breton Language. In Glanville Price (ed.), *The Celtic connection*, 245–275. Buckinghamshire: Rowman & Littlefield.

Isaac, Graham R. 1993. Issues in the reconstruction and analysis of Insular Celtic syntax and phonology. *Ériu* 44. 1–32.

Jacob, Francois. 1977. Darwinism reconsidered. *Le Monde, Translated in WALS Atlas (January 1976)*.

Jäger, Gerhard. 2013. Phylogenetic inference from word lists using weighted alignment with empirically determined weights. *Language Dynamics and Change* 3(2). 245–291. DOI: 10.1163/22105832-13030204.

Jeffers, Robert J. 1976. Syntactic change and syntactic reconstruction. In William M. Christie (ed.), *Current progress in historical linguistics: Proceedings of the second international conference on historical linguistics, Tucson, Arizona 12–16 January 1976*, 1–16. Amsterdam: North Holland.

Jongeling, Karel. 2000. *Comparing Welsh and Hebrew*. Leiden: Center for Non-Western Studies.

Jouitteau, Mélanie. 2005. *La syntaxe comparée du Breton*. Université de Nantes. (Doctoral dissertation).

Kayne, Richard S. 2005. Antisymmetry and Japanese. In Richard S. Kayne (ed.), *Movement and silence*, 215–40. Oxford: Oxford University Press.

Kikusawa, Ritsuko. 2002. *Proto Central Pacific ergativity: Its reconstruction and development in the Fijian, Rotuman and Polynesian languages*. Canberra: Pacific Linguistics.

Kinn, Kari. 2017. Null arguments in Old Norwegian: Interaction between pronouns and the functional categories of the clause. *Linguistic Variation* 17(2). 309–346. DOI: 10.1075/lv.15018.kin.

Koch, John T. 1991. On the prehistory of Brittonic syntax. In James Fife & Erich Poppe (eds.), *Studies in Brythonic word order*, 1–43. Amsterdam: John Benjamins.

Kortlandt, Frederik. 2007. *Italo-Celtic origins and prehistoric development of the Irish language*. Leiden: Editions Rodopi BV.

Krasukhin, Konstantin. 2017. Was Proto-Indo-European an active language? (towards the typology of PIE). *Voprosy Jazykoznanija* (2). Number: 2, 49–76. DOI: 10.31857/S0373658X0000966-1.

Kroch, Anthony, Ann Taylor & Donald Ringe. 2000. The Middle English verb-second constraint. *Textual parameters in older languages*. 353–392. DOI: 10 . 1075/cilt.195.17kro.

Kuhn, A. 1853. Αἰών, aevum, aivs. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete des Deutschen, Griechischen und Lateinischen* 2(3H). 232–235.

Labov, William. 1981. Resolving the neogrammarian controversy. *Language* 57(2). 267–308. DOI: 10.2307/413692.

Le Roux, Pierre. 1957. *Le verbe Breton*. Rennes: J. Plihon.

Ledgeway, Adam. 2012. *From Latin to Romance*. Oxford: Oxford University Press.

Ledgeway, Adam & Ian Roberts (eds.). 2017. *The Cambridge handbook of historical syntax*. Cambridge: Cambridge University Press. DOI: 10.1017/9781107279070.

Ledgeway, Adam & Nigel Vincent. 2025. Latin and Romance: Language contact from a typological perspective. In U. Hoinkes, J. Müller-Lancé & R. Schöntag (eds.), *Manual of historical contact linguistics in romance*. Berlin: de Gruyter.

Lehmann, Winfred P. 1974. *Proto-Indo-European syntax*. Austin: University of Texas Press.

Lehmann, Winfred P. 1993. *Theoretical bases of Indo-European linguistics*. London: Routledge.

Lewis, Henry. 1956. *Llawlyfr Cernyweg Canol*. Cardiff: Cardiff University Press.

Lewis, Henry & Holger Pedersen. 1961. *A concise comparative Celtic grammar*. Göttingen: Vandenhoeck & Ruprecht.

Lewis, Henry & J. R. F. Piette. 1966. *Llawlyfr Llydaweg Canol*. Cardiff: University of Wales Press.

Lightfoot, David W. 1980. On reconstructing a proto-syntax. In P. Ramat (ed.), *Linguistic reconstruction and Indo-European syntax*, 27–45. Amsterdam: John Benjamins.

Lightfoot, David W. 1999. *The development of language: Acquisition, change, and evolution*. Oxford: Blackwell.

Lightfoot, David W. 2002. Myths and the prehistory of grammars. *Journal of Linguistics* 38(1). 113–136. DOI: 10.1017/S0022226701001268.

List, Johann-Mattis, Michael Cysouw & R. Forkel. 2016. Concepticon: A resource for the linking of concept lists. In *Proceedings of the tenth international conference on language resources and evaluation LREC16*, 2393–2400. Marseille: LREC Proceedings.

List, Johann-Mattis, Simon J. Greenhill & Russell D. Gray. 2017. The potential of automatic word comparison for historical linguistics. *PLOS ONE* 12(1). e0170046. DOI: 10.1371/journal.pone.0170046.

Longobardi, Giuseppe. 2017. Principles, parameters and schemata: A radically underspecified UG. *Linguistic Analysis* 41. 517–557.

Longobardi, Giuseppe & Cristina Guardiano. 2009. Evidence for syntax as a signal of historical relatedness. *Lingua* 119(11). 1679–1706. DOI: 10.1016/j.lingua.2008.09.012.

Longobardi, Giuseppe, Cristina Guardiano, Giuseppina Silvestri, Alessio Boattini & Andrea Ceolin. 2013. Toward a syntactic phylogeny of modern Indo-European languages. *Journal of Historical Linguistics* 3(1). 122–152. DOI: 10.1075/jhl.3.1.07lon.

Longobardi, Giuseppe & Alessandro Treves. 2023. Grammatical parameters from a gene-like code to self-organizing attractors. *arXiv preprint arXiv:2307.03152*. DOI: https://doi.org/10.48550/arXiv.2307.03152.

MacCana, Proinsias. 1991. Further notes on constituent order in Welsh. In James Fife & Erich Poppe (eds.), *Studies in Brythonic word order*, 45–80. Amsterdam: John Benjamins.

McCloskey, James. 1990. Resumptive pronouns, A'-binding, and levels of representation in Irish. In Randall Hendrick (ed.), *The syntax of the modern celtic languages*. 199-248. DOI: 10.1163/9789004373228_008.

McCloskey, James. 1996. On the scope of verb movement in Irish. *Natural Language and Linguistic Theory* 14(1). 47–104.

McMahon, April & Robert McMahon. 2005. *Language Classification by Numbers*. Oxford: Oxford University Press.

Meelen, Marieke. 2016. *Why Jesus and Job spoke bad Welsh: The origin and distribution of V2 orders in Middle Welsh*. Utrecht: LOT.

Meelen, Marieke. 2017. Object-initial word order in Middle Welsh narrative prose. In Erich Poppe, Karin Stüber & Paul Widmer (eds.), *Referential properties and their impact on the syntax of Insular Celtic languages*, 145–178. Münster: Nodus Publikationen.

Meelen, Marieke, Nathan W. Hill & Hannes Fellner. 2022. What are cognates? *Papers in Historical Phonology* 7. 44–80. DOI: 10.2218/pihph.7.2022.7405.

Meelen, Marieke & Silva Nurmio. 2020. Adjectival agreement in Middle and early Modern Welsh native and translated prose. *Journal of Celtic Linguistics* 21(1). 1–28. DOI: 10.16922/jcl.21.2.

Meid, Wolfgang. 1968. *Indogermanisch und keltisch*. Innsbruck: Sprachwissenschaftliches Institut der Universität Innsbruck.

Meillet, Antoine. 1954. *La méthode comparative en linguistique historique*. 6th edn. Paris: Champion.

Miller, D. Gary. 1975. Indo-European: VSO, SOV, SVO, or all three? *Lingua* 37. 31–52. DOI: 10.1016/0024-3841(75)90003-0.

Morrison, David A. 2016. Genealogies: Pedigrees and phylogenies are reticulating networks not just divergent trees. *Evolutionary Biology* 43. 456–473. DOI: 10.1007/s11692-016-9376-5.

Neureiter, Nico, Peter Ranacher, Nour Efrat-Kowalsky, Gereon Kaiping, Robert Weibel, Paul Widmer & Remco Bouckaert. 2022. Detecting contact in language trees: A Bayesian phylogenetic model with horizontal transfer. *Humanities and Social Sciences Communications* 9(1). 1–14. DOI: 10.1057/s41599-022-01211-7.

Newmeyer, Frederick J. 2005. Against a parameter-setting approach to typological variation. *Linguistic variation yearbook* 4(1). 181–234. DOI: 10.1075/livy.4.06new.

Newton, Glenda Elizabeth. 2006. *The development and loss of the Old Irish double system of verbal inflection*. University of Cambridge. (Doctoral dissertation).

Nichols, Johanna. 1992. *Linguistic diversity in space and Time*. Chicago: University of Chicago Press.

Niyogi, Partha & Robert C. Berwick. 2009. The proper treatment of language acquisition and change in a population setting. *Proceedings of the National Academy of Sciences* 106(25). 10124–10129. DOI: 10.1073/pnas.0903993106.

Nolan, Brian. 2018. The encoding of negation in Modern Irish: Negation at the layered structure of the clause and noun phrase. In Rolf Kailuweit, Lisann Künkel & Eva Staudiger (eds.), *Applying and expanding role and reference grammar*, 45–70. Freiburg: Freiburg Institute for Advanced Studies.

Ohala, John J. 1981. The listener as a source of sound change. In Carrie S. Masek, Roberta A. Hendrick & Mary Frances Miller (eds.), *Papers from the parasession on language and behavior*, 178–203. Chicago: Chicago Linguistics Society.

Osthoff, Hermann & Karl Brugmann. 1878. *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen*, vol. 1. Leipzig: S. Hirzel.

Ouhalla, Jamal. 1993. Subject-extraction, negation and the anti-agreement effect. *Natural Language and Linguistic Theory* 11(3). 477–518. DOI: 10.1007/bf00993167.

Pacchiarotti, Sara. 2020. On the origins of the ergative marker *wã* in the Viceitic language of the Chibchan family. In Jóhanna Barðdal, Spike Gildea & Eugenio Lujan (eds.), *Reconstructing syntax*, 241–288. Leiden: Brill. DOI: 10.1163/9789004392007_007.

Padel, Oliver J. 2021. English influence on the Middle Cornish verbal system. *Zeitschrift für celtische Philologie* 86(1). 255–296. DOI: 10.1515/zcph-2021-0011.

Panagiotidis, Phoevos. 2015. *Categorial features*. Cambridge: Cambridge University Press.

Pedersen, Holger. 1907. Neues und nachträgliches. *Zeitschrift für vergleichende Sprachforschung* 40. 129–217.

Pellard, Thomas, Robin Ryder & Guillaume Jacques. Forthcoming. The family tree model. In Adam Ledgeway, Edith Aldridge, Anne Breitbarth, Katalin Kiss, Joseph Salmons & Alexandra Simonenko (eds.), *Wiley Blackwell companion to diachronic linguistics*. Oxford: Wiley Blackwell.

Pires, Acrisio & Sarah G. Thomason. 2008. How much syntactic reconstruction is possible? In Gisella Ferraresi & Maria Goldbach (eds.), *Principles of syntactic reconstruction*. Amsterdam: John Benjamins. DOI: 10.1075/cilt.302.04pir.

Poppe, Erich. 1995. Negation in Welsh and 'Jespersen's Cycle'. *Journal of Celtic Linguistics* 4. 99–107.

Press, Ian. 1986. *A grammar of modern Breton*. Berlin & New York: Mouton de Gruyter. DOI: 10.1515/84975.

Press, Ian & Herve Ar Bihan. 2004. *Colloquial Breton: The complete course for beginners*. London: Routledge.

Ram-Prasad, Krishnan & Marieke Meelen. 2022. *Syntactic parameters and the linguistic geography of Indo-European*. Paper presented at the Linguistics Association of Great Britain Annual Meeting.

Rama, Taraka. 2018. Three tree priors and five datasets: A study of Indo-European phylogenetics. *Language Dynamics and Change* 8(2). 182–218. DOI: 10.1163/22105832-00802005.

Ringe, Don, Tandy Warnow & Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100(1). 59–129. DOI: 10.1111/1467-968X.00091.

Roberts, Ian. 1997. Restructuring, head movement, and locality. *Linguistic Inquiry* 28. 423–460.

Roberts, Ian. 1998. Review of historical syntax in cross-linguistic perspective. *Romance Philology* 51(1995). 363–370. DOI: 10.1017/s0022226798267119.

Roberts, Ian. 2007. *Diachronic syntax*. Oxford: Oxford University Press.

Roberts, Ian. 2010. *Agreement and head movement: Clitics, incorporation, and defective goals*. Cambridge: MIT Press.

Roberts, Ian. 2019. *Parameter hierarchies and Universal Grammar*. Oxford: Oxford University Press. DOI: 10.1093/oso/9780198804635.001.0001.

Roberts, Ian, James Baker, Marieke Meelen, Krishanan Ram-Prasad & Elena Isolani. 2022. Extending parametric comparison: Some preliminary results in older Indo-European languages. In *Diachronic generative syntax*.

Roberts, Ian & Anna Roussou. 2003. *Syntactic change: A minimalist approach to grammaticalization*. Cambridge: Cambridge University Press.

Ross, Malcolm. 2015. The argument indexing of early Austronesian verbs. In *Historical linguistics 2013: Selected papers from the 21st International Conference on Historical Linguistics, Oslo, 5-9 august 2013*, vol. 334, 257–79. Amsterdam: John Benjamins.

Rumsey, Alan. 1987. The chimera of Proto-Indo-European ergativity: Lessons for historical syntax. *Lingua* 71. 297–318.

Schleicher, August. 1853. 'Die ersten Spaltungen des indogermanischen Urvolkes. *Allgemeine Monatsschrift für Wissenschaft und Literature* 1853. 786–7.

Schrijver, Peter. 1995. *Studies in British Celtic historical phonology*, vol. 5. Rodopi.

Schrijver, Peter. 1997. *Studies in the history of Celtic pronouns and particles*. Dublin: Department of Old Irish, National University of Ireland.

Seržant, Ilja A. 2015. An approach to syntactic reconstruction. In Carlotta Viti (ed.), *Perspectives on historical syntax*, 117–154. Amsterdam: John Benjamins. DOI: 10.1075/slcs.169.05ser.

Sheehan, Michelle. 2014. Portuguese, Russian and the theory of control. In Hsin-Lun Huang, Ethan Poole & Amanda Rysling (eds.), *Proceedings of the 43rd annual meeting of the North East Linguistic Society*, vol. 43(2), 115–26.

Shlonsky, Ur. 1992. Resumptive pronouns as a last resort. *Linguistic Inquiry* 23. 443–68.

Sims-Williams, Patrick. 2003. *The Celtic inscriptions of Britain: Phonology and chronology, c. 400-1200*. Boston: Blackwell.

Smith, Jennifer. 2000. *Synchrony and diachrony in the evolution of English: Evidence from Scotland*. University of York. (Doctoral dissertation).

Smitherman, Thomas. 2015. Reconstructing non-canonical argument structure for Proto-Indo-European: Methodological questions and progress. In Carlotta Viti (ed.), *Studies in language companion series*, vol. 169, 93–116. Amsterdam: John Benjamins. DOI: 10.1075/slcs.169.04smi.

Stump, Gregory T. 1984. Agreement vs incorporation in Breton. *Natural Language and Linguistic Theory*. 289–348. DOI: 10.1007/BF00133790.

Sutrop, Urmas. 2012. Estonian traces in the tree of life concept and in the language family tree theory. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 3(1). 297–326. DOI: 10.12697/jeful.2012.3.1.15.

Szemerényi, Oswald. 1996. *Introduction to Indo-European linguistics*. Oxford: Clarendon Press.

Thomason, Sarah G. 2007. Language contact and deliberate change. *Journal of language contact* 1(1). 41–62. DOI: 10.1163/000000007792548387.

Thurneysen, Rudolf. 1885. Oskische einschubsvocale. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der Indogermanischen Sprachen* 27(2. H). 181–182.

Thurneysen, Rudolf. 1927. Altirische grammatik. *Indogermanische Forschungen* 44(s1). 370–372.

Toman, Jindrich. 1998. A discussion of resumptives in colloquial Czech. In Z. Boskovic (ed.), *Formal approaches to slavic linguistics*, vol. 6, 303–318. Ann Arbor: Michigan Slavic Publications.

Toorians, Lauran. 2014. *Towards a grammar of Middle Cornish*. http : / / laurantoorians.com/?page_id=128 (5 June, 2025).

Trask, R. L. 1996. *Historical linguistics*. London: Arnold.

Traugott, Elizabeth Closs & Graeme Trousdale. 2013. *Constructionalization and constructional changes*. First edition (Oxford linguistics 6). Oxford: Oxford University Press.

Uhlenbeck, C. Cornelis. 1901. Agens und Patiens im Kasussystem der indogermanischen Sprachen. *Indogermanische Forschungen* 12. 170–171.

Vaillant, André. 1936. L'ergatif indo-européen. *Bulletin de la Société de Linguistique de Paris* 37. 93–108.

Van der Wal, Jenneke. 2020. From microparameters to macroparameters: A Bantu case study. In Theresa Biberauer, Sten Vikner, Andras Bárány & Jamie Douglas (eds.), *Syntactic architecture and its consequences I: Syntax inside the grammar*. Language Science Press. DOI: 10.5281/zenodo.4041229.

Van Gelderen, Elly. 2011. *The linguistic cycle: Language change and the language faculty*. Oxford: Oxford University Press.

Van Gelderen, Elly. 2022. *Third factors in language variation and change*. Cambridge: Cambridge University Press.

Villar, Francisco. 1983. *Ergatividad, acusatividad y género en la familia lingüística indoeuropea*. Salamanca: Universidad de Salamanca.

Viti, Carlotta. 2015. Historical syntax: Problems, materials, methods, hypotheses. In Carlotta Viti (ed.), *Perspectives on historical syntax*, 1–32. Amsterdam: John Benjamins. DOI: 10.1075/slcs.169.01vit.

Von Mengden, Ferdinand. 2008. Reconstructing complex structures: A typological perspective. In Gisella Ferraresi & Maria Goldbach (eds.), *Principles of syntactic reconstruction*. Amsterdam: John Benjamins. DOI: 10.1075/cilt.302.06men.

Wackernagel, Jacob. 1892. Über ein Gesetz der Indogermanischen Wortstellung. *Indogermanische Forschungen* 1. 333–436.
Walkden, George. 2009. *The comparative method in syntactic reconstruction*. University of Cambridge. (MA thesis).
Walkden, George. 2011. Abduction or inertia? The logic of syntactic change. In Chris Cummins, Chi-Hé Elder, Thomas Godard, Morgan Macleod, Elaine Schmidt & George Walkden (eds.), *Proceedings of the Sixth Cambridge Postgraduate Conference in Language Research*, 230–239. Cambridge Institute of Language Research.
Walkden, George. 2014. *Syntactic reconstruction and Proto-Germanic*. Oxford: Oxford University Press.
Warner, Anthony R. 1993. *English auxiliaries: Structure and history*. Cambridge: Cambridge University Press.
Watkins, Calvert. 1964. Preliminaries to the reconstruction of Indo-European sentence structure. In Horace G. Lunt (ed.), *Proceedings of the 9th International Congress of Linguists*, 1035–1045. The Hague: Mouton.
Watkins, Calvert. 1966. The origin of the f-future. *Ériu* 20. 67–81.
Watkins, Calvert. 1976. Towards Proto-Indo-European syntax: Problems and pseudo-problems. In Stanford Steever, Carol Walker & Salikoko Mufwene (eds.), *Papers from the parasession on diachronic syntax*, 306–326. Chicago: Chicago Linguistic Society.
Watkins, Calvert. 1995. *How to kill a dragon: Aspects of Indo-European poetics*. Oxford: Oxford University Press.
Weiss, Michael. 2022. Italo-Celtic. In Thomas Olander (ed.), *The Indo-European language family*, 102–113. Cambridge: Cambridge University Press. DOI: 10.1017/9781108758666.007.
Weiss, Michael. Forthcoming. Italo-Celtic preverbs and adpositions. In Joseph F. Eska, Silva Nurmio, Peadar Ó Muircheartaigh & Paul Russell (eds.), *The Palgrave handbook of Celtic languages and linguistics*. London: Palgrave Macmillan.
Williams, Nicholas J. A. 2011. Middle and late Cornish. In Elmar Ternes (ed.), *Brythonic Celtic - Britannisches Keltisch: From Medieval British to Modern Breton*, 237–357. Bremen: Hempen.
Willis, David. 1998. *Syntactic change in Welsh: A study of the loss of the verb-second*. Oxford: Oxford University Press.
Willis, David. 2000. On the distribution of resumptive pronouns and *wh*-trace in Welsh. *Journal of Linguistics* 36(3). 531–573.
Willis, David. 2006. Negation in Middle Welsh. *Studia Celtica* 40(1). 63–88.

Willis, David. 2011. Reconstructing last week's weather: Syntactic reconstruction and Brythonic free relatives. *Journal of Linguistics* 47(2). 407–446. DOI: 10.1017/S0022226710000381.

Willis, David. 2013. The history of negation in the Brythonic Celtic languages. In David Willis, Christopher Lucas & Anne Breitbarth (eds.), *The history of negation in the languages of Europe and the Mediterranean, volume 1: Case studies*, 239–298. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199602537.003.0007.

Willis, David. Forthcoming. Comparative Brittonic syntax. In Joseph F. Eska, Silva Nurmio, Peadar Ó Muircheartaigh & Paul Russell (eds.), *The Palgrave handbook of Celtic languages and linguistics*. London: Palgrave Macmillan.

Wmffre, Iwan. 1998. *Central Breton*. Munich: Lincom.

Part IV

Sociolinguistics and language revitalization

Chapter 13

Guth Thormoid: The "Island Voice" of Norman Maclean

Gordon Wells
University of the Highlands and Islands

This chapter samples and contextualises some of the multi-faceted Gaelic contributions by the multi-talented creative icon, Norman Maclean, to the *Guthan nan Eilean* 'Island Voices' online language capture and curation project. These include, in particular, Norman's final *Saoghal Thormoid* 'Norman's World' series of videoed conversations, recorded in April 2016 in which he spoke reflectively of his memories and impressions of bilingual life in Glasgow and the Hebrides from the middle of the Twentieth Century onwards. In addition to offering a vivid first-person voiced and experiential account of Gaelic life over a tumultuous period for the language, the Island Voices adherence to basic linguistic principles pays dividends in relation to some initially unpredicted spin-off applications, with potential for further development.

1 Introduction

In this chapter, I describe the history and content of a documentation exercise focused on a singular individual, whose work emerged organically out of the Hebridean *Guthan nan Eilean* 'Island Voices' language capture and curation project (Wells 2023). Specifically, it relates to the output of Norman Maclean, who was a person of many different talents and skills, and whose linguistic capacity was remarkable. As a case study, the approach is necessarily selective rather than comprehensive, seeking to give an indicative representation of Norman's various contributions, though with a particular focus on his final recordings for the project.

Gordon Wells. 2025. *Guth thormoid*: the "island voice" of norman maclean. In Andrew Carnie, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond & Luis Irizarry (eds.), *Foundational approaches to Celtic linguistics*, 471–493. Berlin: Language Science Press. DOI: 10.5281/zenodo.15654881

I look at this output largely through applied linguistic eyes, with a fundamental orientation towards throwing light on the everyday Gaelic speech and culture of ordinary people. This may have the appearance of a somewhat paradoxical exercise. There is an inherent danger in attempting to draw widely applicable lessons from the study of exceptional exponents of the language. Nevertheless, I hope to demonstrate that this exercise has delivered a valuable resource worthy of close study by anyone with an interest in contemporary vernacular Gaelic.

As a computer-mediated project with a focus on recorded Gaelic speech, its description through written English presents a challenge. However, numerous online links to recordings are given in footnotes.

I start with Norman's self-introduction in *Saoghal Thormoid* 'Norman's World,' his final and weightiest contribution to Island Voices:

> *Tha fhios aig a h-uile Gàidheal cò e. M' athair, 's e Niall Mòr, mac Iain Eòghainn Ruaidh à Cillmoluaig, Tiriodh. Maclean, a h-uile dàrna fear ann an Tiriodh 's e Maclean a th' orra. Ach sin mise air taobh m' athar. An gille aig Niall Mòr, mac Iain Eòghainn Ruaidh. Air taobh mo mhàthar, mo mhàthair Peigi Bheag, nighean Anna Bige. Nis 's e Anna Bheag a bha seo mo sheanmhair, 's i nighean Aonghais 'ic Iain Mhòir à Hàcleit.*
>
> 'Every Gael knows who he is. My father, he's Big Niall, son of Iain Eòghainn Ruaidh from Kilmaluaig, Tiree. Maclean, every second person on Tiree is a Maclean. But that's me on my father's side. Son of Big Niall, son of Iain Eòghainn Ruaidh. On my mother's side, Wee Peggy, daughter of Wee Ann. Now my granny Wee Ann, she's the daughter of Angus, son of Big Iain from Hacklett.' (Tormod MacGill-Eain)[1]

This can be taken as an example of a once typical Gaelic personal introduction, through the subject's *sloinneadh*, or genealogy. It is a traditional naming system which relies largely on patronymics, though matronymics may also be used. In fact, in this case we're given both sides of Norman's family.

By way of my own self-introduction, I also have a *sloinneadh*: *Is mi an gille as òige aig Anna Sheonaidh 'ic Ghilleasbaig* 'I'm the youngest son of Anna, daughter of Seonaidh, son of Gilleasbaig.' It is an important introductory point to make that this is how people in the North Uist community into which my mother was born know themselves and know each other. This is how you identify friends, relations, and other community members. In other terms it might be so said to help

[1]https://youtu.be/nCR2mVnDcX0

present or mediate, at least to some extent, my own "positionality" (Greenbank 2003) with respect to the community.

That being said, I was actually born in Edinburgh and raised speaking mostly English, firstly for a short while in India, then in England, and did not "come home" to live in Uist until my fortieth year. In this respect, my perspective on the islands community in which I then became a resident remains in part an external one, complemented by an academic background in Linguistics and Applied Linguistics, and a professional training and working experience as a language teacher and researcher.

The *Guth Thormoid* in the title of this article translates as 'Norman's Voice'. The "Island Voice" element is placed in quotation marks for emphasis, because Norman had many voices, and they weren't all island ones by any means. However, this article concentrates on his contributions to the above-mentioned Island Voices or *Guthan nan Eilean* project, which I co-ordinate.

The rest of this article gives further information about Norman, about the Island Voices project, and about his contributions to it. Attention is paid to his *Saoghal Thormoid* series of conversations. Following this, a brief description is given of some spin-off applications, which were quite unpredicted at the outset of the work, but which can be seen as natural progressions out of the fundamentally linguistic orientation of the project at its inception. The article concludes with some summary points on Norman's special contribution to the Island Voices project.

2 Norman Maclean

Many appreciations have been written about the multi-talented and multi-lingual Norman Maclean (26 December 1936–31 August 2017), or *Tormod MacGill-Eain* as he was known in Gaelic. His creative work was wide-ranging across various genres, with notable success as a comedian, novelist, poet, piper, singer, actor and broadcaster. He famously won both Bardic Crown and Gold Medal at the same Royal National Mòd in 1967. Born in Glasgow to Hebridean parents (from North Uist and Tiree), he was raised in various Gaelic-speaking households in the city itself, as well as Lochaber and then Benbecula when evacuated as a child during the war period. He struggled openly with alcoholism for many years, but he found some peace on retiring to Uist for the final decade or so of his life. His was a well-known and well-loved name in the Gaelic world.

My own personal acquaintance with him was limited. I did not get to know him until he had entered his eighth decade. But I'm happy to vouch for his talent,

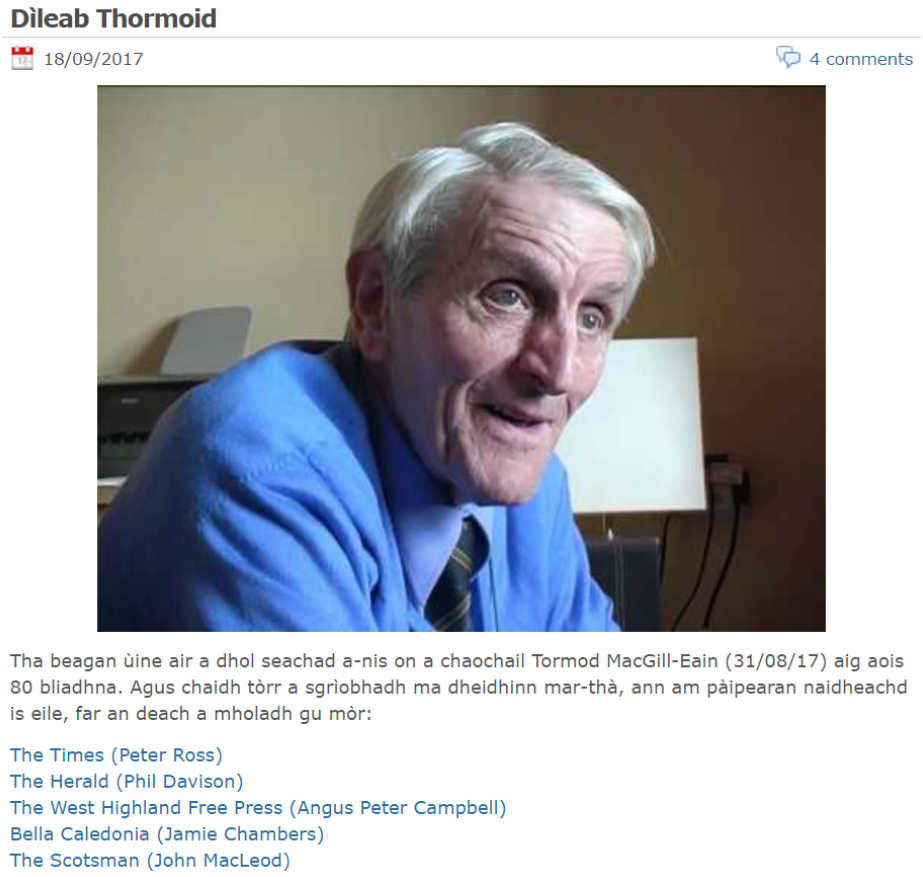

Dìleab Thormoid
18/09/2017
4 comments
Tha beagan ùine air a dhol seachad a-nis on a chaochail Tormod MacGill-Eain (31/08/17) aig aois 80 bliadhna. Agus chaidh tòrr a sgrìobhadh ma dheidhinn mar-thà, ann am pàipearan naidheachd is eile, far an deach a mholadh gu mòr:
The Times (Peter Ross)
The Herald (Phil Davison)
The West Highland Free Press (Angus Peter Campbell)
Bella Caledonia (Jamie Chambers)
The Scotsman (John MacLeod)

Figure 1: *Dìleab Thormoid* 'Norman's Legacy' screenshot: https://guthan.wordpress.com/2017/09/18/dileab-thormoid-dhuinn/

which I was pleased to witness on various occasions, and for his generosity, particularly with his time, given the signal contribution he made in various ways to the Island Voices project itself. So when he passed we did write our own tribute (see Figure 1). It is written in Gaelic, but the top paragraph actually links through to a number of obituaries which were published in national and local press as well as online, so further information on his life and times is readily available in English through the Island Voices site.

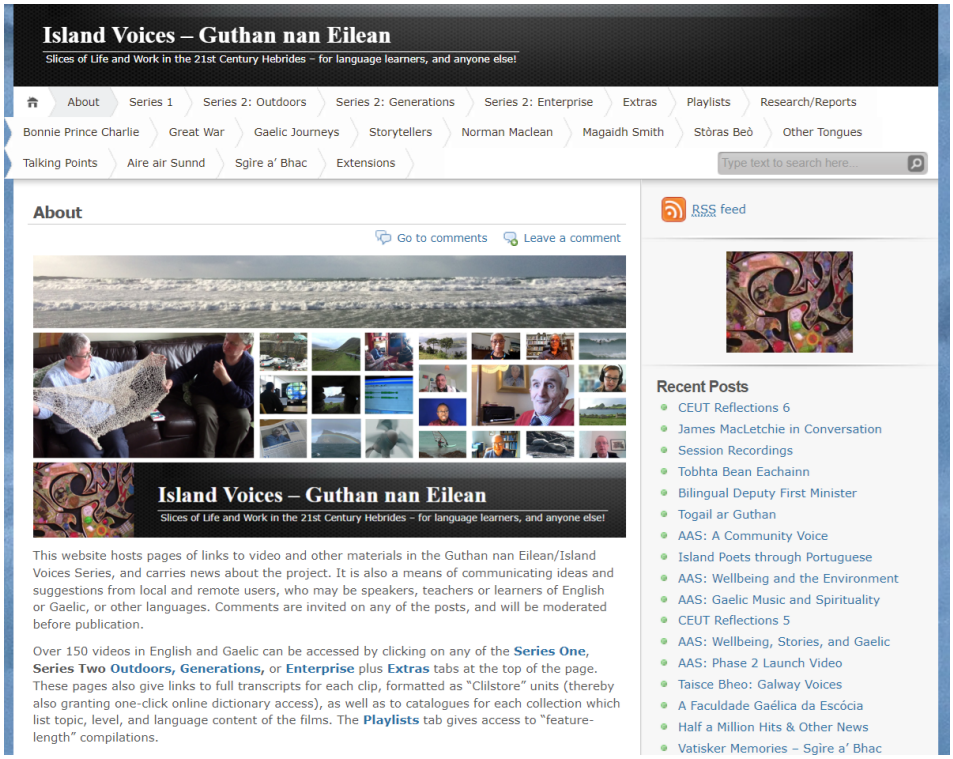

Island Voices – Guthan nan Eilean
Slices of Life and Work in the 21st Century Hebrides – for language learners, and anyone else!
About
Series 1
Series 2: Outdoors
Series 2: Generations
Series 2: Enterprise
Extras
Playlists
Research/Reports
Bonnie Prince Charlie
Great War
Gaelic Journeys
Storytellers
Norman Maclean
Magaidh Smith
Stòras Beò
Other Tongues
Talking Points
Aire air Sunnd
Sgìre a' Bhac
Extensions
Type text to search here...
About
Go to comments
Leave a comment
RSS feed
Recent Posts
CEUT Reflections 6
James MacLetchie in Conversation
Session Recordings
Tobhta Bean Eachainn
Bilingual Deputy First Minister
Togail ar Guthan
AAS: A Community Voice
Island Poets through Portuguese
AAS: Wellbeing and the Environment
AAS: Gaelic Music and Spirituality
CEUT Reflections 5
AAS: Wellbeing, Stories, and Gaelic
AAS: Phase 2 Launch Video
Taisce Bheo: Galway Voices
A Faculdade Gaélica da Escócia
Half a Million Hits & Other News
Vatisker Memories – Sgìre a' Bhac
Island Voices – Guthan nan Eilean
Slices of Life and Work in the 21st Century Hebrides – for language learners, and anyone else!
This website hosts pages of links to video and other materials in the Guthan nan Eilean/Island Voices Series, and carries news about the project. It is also a means of communicating ideas and suggestions from local and remote users, who may be speakers, teachers or learners of English or Gaelic, or other languages. Comments are invited on any of the posts, and will be moderated before publication.
Over 150 videos in English and Gaelic can be accessed by clicking on any of the Series One, Series Two Outdoors, Generations, or Enterprise plus Extras tabs at the top of the page. These pages also give links to full transcripts for each clip, formatted as "Clilstore" units (thereby also granting one-click online dictionary access), as well as to catalogues for each collection which list topic, level, and language content of the films. The Playlists tab gives access to "feature-length" compilations.

Figure 2: About Island Voices Screenshot: https://guthan.wordpress.com/about/

3 Island Voices

Island Voices has a presence in various corners of the Internet, (principally on http://WordPress.com and http://youtube.com). It also has social media channels through Twitter and Facebook. The project is conceived as a "language capture and curation project." We record linguistic material, and we display it. The approach has been bilingual from the start. It is now multilingual, but Gaelic has a central role given the Hebridean locus and focus of the project. It is all online, and there's free and open access across the world – not just in the islands which are the geographical centre of interest. We're open to everybody.

The project has gone through various phases, starting in 2005 and going online three or four years after that. Initially, the orientation was towards pedagogical development in terms of language learning and teaching – looking at methods and methodologies, and technologies. That was a staff-led period, in which

project staff – principally myself – produced short documentaries to focus on a particular topic or venue in the community, and complemented those with authentic speech (Widdowson 1998) talking-head interviews with local members of the community who had a particular connection with that topic.

After establishing a central core of materials (Series One and Two), we moved on from the initial staff-led approach to something more participatory, encouraging local community members and local community groups to initiate the same kind of work – to collect or create recordings and then share them through our Island Voices websites. And that continues to be the ethos as we go on. The tabs at the top of Figure 2 illustrate how over the years we've created a good number of different pages, which show the topical range of the full Island Voices project. A page basically consists of one or two tables, and each table contains multiple links to materials which we've produced in various formats. Among the collection of pages on the site there is a dedicated 'Norman Maclean' one, in light of the significant amount of material which he himself produced.

The first 150 videos were created in the staff-led phase referred to above. These are the more pedagogically oriented materials in Series One and Series Two Outdoors, Generations, and Enterprise. Since then, with the broadening out of the focus and participation, we've more than doubled the number of videos which we have on the YouTube site. And the orientation itself has widened from a focus on language learning and language teaching to welcome participation across the community, including in particular from people to whom Gaelic proficiency and habitual use have come as an unremarkable and natural function of being raised in a Hebridean home.

Of course, we're pleased if we can continue to help people learn the language. We think that we can help them do that through profiling unmarked, functional language in use.

The project title "Island Voices" has two elements. Census returns from 2011 explain the significance of the 'Islands' element in relation to Gaelic. It is worth noting that the bar charts in Figure 3, based on the 2011 census, rely on self-report data from questions asked across Scotland about speaking Gaelic (Scotland's Census 2011). Even with this caveat, they give a clear indication that of the various local authorities it is only in the Outer Hebrides that such a significant proportion of the population is found speaking Gaelic, or at least claiming to speak it. It dwarfs the figure for the proportion of the total population of Scotland. So there is clearly a special connection between Gaelic and these islands.[2]

[2] Ó Giollagáin et al. (2020) offer fine-grained analysis to suggest a figure of 11,000 vernacular speakers of Gaelic in their islands study area.

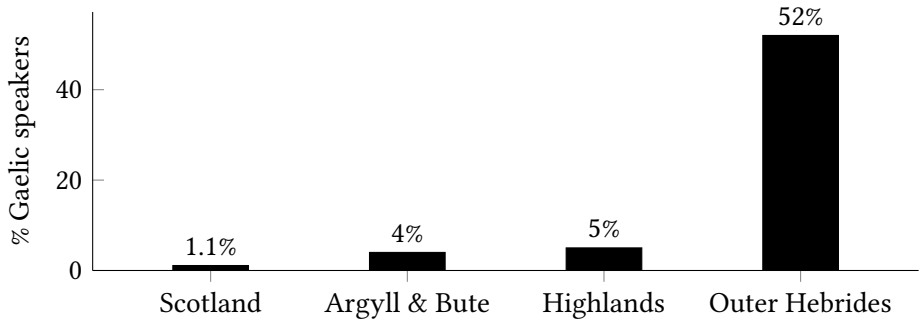


Figure 3: 2011 Census returns on Gaelic speaker percentages (Scotland's Census 2011)

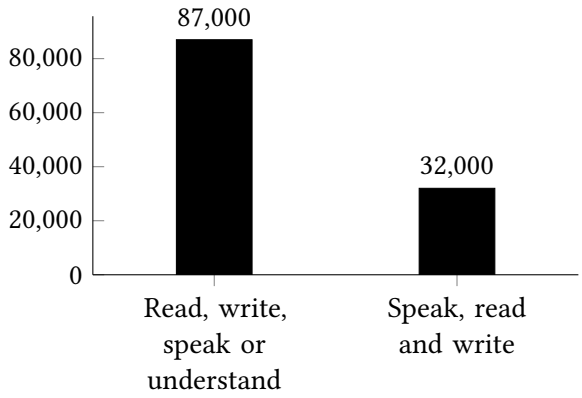


Figure 4: Census figures on Gaelic reading, writing, and speaking skills (Scotland's Census 2011)

Census figures, as shown in Figure 4, on Gaelic reading, writing, and speaking skills also underline the significance of the "Voices" element in the project title. There is a clear discrepancy between the number of people who speak the language and the number who read and write it as well. Island Voices is about oracy, and the primacy of speech (Sapir 1921, Bloomfield 1933, de Saussure 1966, Chafe & Tannen 1987, Wells 2020). That is what we want to address and profile particularly. It is a distinguishing feature of the overall picture that the best speakers of Gaelic, being the people who have been brought up using it, have not necessarily received an education through it. They are certainly literate, but their trained literacy is in their other language. We should not let them feel that their Gaelic skills are not recognised or valued simply because they do not read or write it. We need to actively profile the spoken language, and that is what Island Voices undertakes to do as a primary goal.

4 Norman Maclean and Island Voices

Figure 5 collects a number of screenshots that trace the history of Norman's involvement with Island Voices. "Island Voices Enterprise" is an example from the staff-led beginnings of the project, where we had topics that were introduced in staff-produced short documentaries, complemented with talking head interviews with people associated with the topic. Below this, the table from the Island Voices Enterprise page tab shows the bilingual presentation – English on the left, Gaelic on the right – with a short documentary on the *Am Pàipear* community newspaper in plainer "teacher talk" language, followed by the personal interviews, which are the interesting part from the point of view of recorded authentic speech. Tormod MacGill-Eain, or Norman Maclean, is listed amongst these. This dates from 2010.

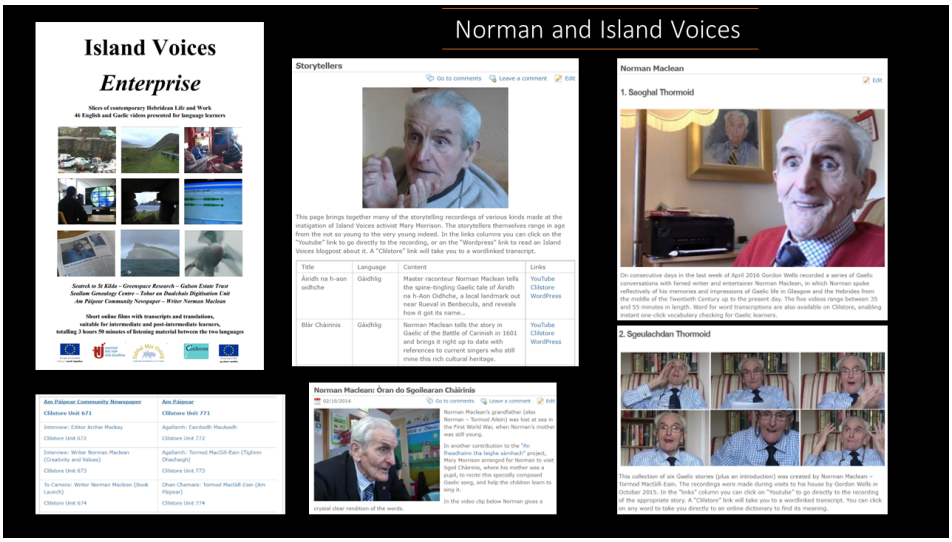


Figure 5: Norman and Island Voices Screenshot Collage (click on link for larger image): https://guthan.files.wordpress.com/2023/06/tormod-slide-8.pdf

In the middle section of Figure 5, we move into more participatory work involving local individuals and groups. As an example, we created a page called Storytellers, because one of our Island Voices activists, Mary Morrison, was keen to go out into the community and record people telling stories. We have two of Norman's stories at the top of that table. This moves us into a different functional genre – that of narration. The second example in the central column is from our Great War project, which was led by the North Uist Historical Society in 2014,

marking the 100th anniversary of the outbreak of World War One. Local society members made recordings in the community. One of these was a video of Norman reciting one of his own poetic compositions to local schoolchildren in the Carinish School where his mother had been a pupil.

On the right of Figure 5 is Norman Maclean's own dedicated page. It has two parts to it – Number Two, *Sgeulachdan Thormoid*, and Number One, *Saoghal Thormoid*. *Sgeulachdan* means 'stories'. This was his initiative. He came to us, and although it is listed as Number Two, it was in fact the first of these two parts chronologically. Norman wanted to record some of his short stories which he had been writing for the benefit essentially of young people who might want to get up on stage and perform a recitation in a local competition or even at the National Mòd. He was keen to get these recorded and available for people to view, so we were more than happy to accommodate his wish.

That sparked *Saoghal Thormoid* very soon thereafter when the then director of Soillse, Professor Conchúr Ó Giollagáin, spotted *Sgeulachdan* and suggested we follow it up with some discursive recordings in which Norman would be encouraged to look back on his long and interesting life. Luckily, Norman was happy to oblige, and I was given the privilege of going into his house on consecutive days for a full week, just to let him speak and to record his words. This substantial and significant *Saoghal Thormoid* collection was Norman's last contribution to Island Voices. The videos were placed online with Clilstore transcriptions (see Section 7.1 below) on completion in November 2016. A printable volume was later uploaded (Wells 2017), including a foreword from Ó Giollagáin concluding with this summation (p4): "Tormod's life is an acknowledgement of the cultural wealth of Gaelic society, and by virtue of this archive, he represents an ambassador to its future."

The following account gives a summary description of each day's conversation supplemented with brief commentary and selected quotes to highlight points of particular interest.

5 Saoghal Thormoid: Norman's world

5.1 Monday: Ancestry

Summary:"Every Gael knows who he is." Norman talks about his genealogy, on both sides of the family, and how these family networks played an important part in his early upbringing in Glasgow, Lochaber, and Benbecula. He has clear memories of his paternal grandfather teaching him songs, a man who himself

won a prize for Gaelic singing at the Falkirk Tryst of 1878. His maternal grandmother, meanwhile, migrated to Glasgow from North Uist and never learned to speak English, functioning socially just within the Gaelic-speaking community of Glasgow of that time. Norman reflects on how community relations were experienced from different perspectives in his childhood.

In giving Norman's *sloinneadh* in the introduction to this article, I've already quoted from this conversation. But in addition to the importance of the *sloinneadh* (which might contain other elements beyond the purely genealogical – for example physical attributes, such as *Niall Mòr* – 'Big Neil', or *Peigi Bheag* – 'Wee Peggy'), the reference to Gaelic signals a key theme, and the introduction of social commentary, right from the start, is also noteworthy. It recurs throughout all five days of the conversations. He was constantly contextualising what he was telling us, by explaining what was happening around him at the time, as this extract illustrates:

> "*… Ach, 's e cailleach smart a bh' innte.* But, *thuirt mo mhàthair, an uair ud ann an Glaschu, cha ruigeadh i leas Beurla a bhith aice. Bha uiread de Ghàidheil timcheall oirre, agus bhiodh i – daoine a' bruidhinn mu dheidhinn* Muslim isolation*is gnothaichean – cha do rinn i oidhirp sam bith Beurla ionnsachadh. Bha nighean aice a sgrìobhadh litrichean dhi – seo mo mhàthair. Chan e gur e* Nobel prize-winner*a bha nam mhàthair, ach,* you know*, bha comas sgrìobhaidh aice. Bhiodh ise a' sgrìobhadh. Bha na gillean, Tormod is Seumas is Cailean – rachadh iadsan dha na bùithtean dhi airson biadh, agus, recreation mar a chanas iad, well rachadh i a thaigh a' Bharraich air* Paisley Road…"
>
> '… But she was a smart old woman. But, my mother said, at that time she didn't need English. There were so many Gaels around her, and she would – people talking about Muslim isolation and things – she made no effort to learn English. She had a daughter who would write letters for her – this is my mother. Not that my mother was a Nobel prize-winner, you know, but she could write. She would write. The boys, Tormod and Seumas and Cailean – they would go to the shops for her for food, and, recreation as they say, well she would go to the Barrach's house on Paisley Road…' (Tormod MacGill-Eain) [3]

Here Norman talks about his grandmother and explains about her Gaelic and English language skills and practices in a description which might equally fit numerous other migrant community circumstances. She did not speak any English

[3] https://youtu.be/nCR2mVnDcX0

as she did not need to. She had children who'd look after everyday communicative needs. And if she needed entertainment she would go and visit other Gaels in Glasgow.

So he keeps a running social commentary, and it is not just about then, but it is also about now, with the parenthetical reference to so-called "Muslim isolation". This recording was made in 2016 around the time when a national report (Casey 2016) was in the news, which affected to be concerned about community integration in the UK, and expressed some worries about levels of English language skills amongst certain communities. Norman clearly had no truck with the apparent controversy. It did not impress him at all that it was considered to be a matter of some concern that when people moved in a community to a new place they continued speaking their own language, particularly if they had other people from their own community around them. He had seen something very similar before in his own family history, and so made the link to contemporary life.

5.2 Tuesday: Education

Summary: After offering some further thoughts on the dominant Catholic/Protestant divide in the Glasgow of his youth, Norman goes on to trace his educational journey, with customary vivid detail and illustrative anecdote, through primary schools in Lochaber, Benbecula and Glasgow, and on to Bellahouston Academy and Glasgow University. He discusses the constraints on, and the opportunities for, varied language choices he and others made in these contexts, within and outwith home and school environments, reflecting also on the Gàidheal-Gall relationship in Glasgow, and some of the wider educational choices he made at that time.

On Tuesday, when Norman describes his journey through education, significantly it is a history not confined to Glasgow. It also includes other schools in the Highlands and in Benbecula during the war period when he was evacuated out of the big city. It is worth noting in passing that this account is full of rattling yarns, narrated in his own unique style. Creamola crops up. So does Louis d'or. The broader point here is that community relations are figuring again. He talks about Glasgow and the very dominant and obvious Catholic/Protestant divide as it was felt in the part where he lived, and the relationship between *Gàidheal* and *Gall*, between Gaelic speaker and non-Gaelic speaker. These were salient points for him as he was progressing through his education.

> *Ach nach eil e neònach mar a bha m' inntinn ag obair nuair a bha mi beag. Nuair a bha mi còmhla ri muinntir Ghlaschu, còmhla ris a' chloinn, cha robh ann*

> *ach dà roghainn san t-saoghal. Cha robh ann ach dà threubh – Pròstanaich is Caitligich. Agus cha robh iad a' tighinn ri chèile uair sam bith. Nis, nach eil e neònach, leis a' chloinn aig an robh pàrantan às Uibhist a Deas, 's à Èirisgeigh 's à Barraigh, Caitligich, cha robh dad sam bith, cha robh cnap-starra sam bith eadar an dà threubh, ach le muinntir Ghlaschu bha e cho soilleir, bha e dìreach ann an dubh 's an geal. Agus, nach eil e neònach, tha am beum a tha seo eatorra nam cheann fhathast.*
>
> 'But isn't it strange how my mind worked when I was small. When I was with Glasgow folk, with the children, there were only two choices in the world. There were only two tribes – Protestant and Catholic. And they never came together. Now, isn't it strange, with the children with parents from South Uist, Eriskay, Barra, Catholics, there was nothing, no obstacle between the two tribes, but with Glasgow folk it was so clear, it was just in black and white. And, isn't it strange, this gap between them is in my head still.' (Tormod MacGill-Eain)[4]

Relating this history in his eighth decade, he's very conscious of the contradictions which made perfect sense to him at the time. So he is open about his own childhood prejudices and about their lasting impact. While he is still a showman, he is at a mature stage now in his life and he is able to openly self-question, which is highly valuable from our point of view as we look to get informed comment from a good observer. The self-questioning element running through his commentary speaks to an ongoing critical and analytical perspective.

5.3 Wednesday: Communities

Summary: Norman describes and reflects upon changes he has witnessed in Gaelic community life over the years, both in Glasgow and in the Hebrides, highlighting some paradoxes and tensions. In former times geographical horizons may have been much closer in comparison with the global awareness and contacts modern connectivity enables, yet the latter may not lead to a sense of greater connectedness. He discusses how, while the Gaelic community in Glasgow may have tended to envisage itself, in his eyes, in a higher or somewhat exclusive position in relation to other Glaswegians, there was nonetheless a strongly felt imperative to acquire their language. Conversely, while young Gaels might be envied by their peers in some ways, they did not feel their language was

[4] https://youtu.be/eiIzDMYAI_A

respected by non-speakers, with apparent racial imprecations sometimes experienced. Lastly, in discussing how broadly the term *Gàidhealach* might be applied, he depicts in more detail the links and fissures, as he saw them, between Glasgow communities of Irish and Scottish Island/Highland extraction.

This is a longer summary, as it was a longer session involving a much deeper dive into some of these community issues. He talks about different kinds of group membership and how they rub along, or rub up against each other, and different kinds of identities and experiences, depending on the categories under consideration. It could be language. It could be religion. It could also be skin colour. It is a wide-ranging discussion, including mention of explicitly racial language. He relates how a slur used against people of South Asian heritage in the UK was turned against Gaelic speakers. Norman and his friends speaking Gaelic to each other in the playground, or other community space, might have this slur applied against them but prefaced with the word 'White.' For example, "Stop that you White ___. Are you talking about me?" This adds another layer to the complex community relations aspect of life that he observed and experienced in Glasgow as he grew up.

The quote below picks out a topical comment relating this complexity to what he was seeing in today's world. He is talking about the boundaries and the horizons that people might feel, and pointing out the differences between Glasgow folk (*muinntir Ghlaschu*) and Gaelic folk (*muinntir Ghàidhealach*) in Glasgow in those days, but highlighting modern changes. (The full connotations of the word *Gàidhealach* are difficult to capture succinctly in English but the base meaning can be broadly construed as 'pertaining to Gaelic/Highland language and culture.')

> *… Ann an Glaschu, cha robh ach pàirt dham bheatha Gàidhealach. An còrr, a' chuid bu mhotha, 's ann am measg muinntir Ghlaschu a bha mi, agus bha mi, mar gum biodh, ga leantail-sa a-muigh air na sràidean. Ach nuair a thiginn a-staigh, 's e muinntir Ghàidhealach a bha a' tighinn a chèilidh oirnn, agus bhithinn ag èisteachd riutha-san… Ach tha atharrachadh mòr air tighinn air an dà chuid… Tha a h-uile duine sgapte bho chèile, agus tha iad, tha iad a' faighinn eòlas air a' chèile tro na h-innealan… Chan eil iad faisg air a' chèile ann. Agus tha an aon rud ann an Glaschu agus tha sna Ceallan shìos. Chan eil daoine – tha iad mar gum biodh air leth bho chèile, ann an Glaschu agus air a' Ghàidhealtachd.*
>
> '… In Glasgow, only part of my life was Gaelic. For the rest, the larger part, I was among Glasgow folk, and I followed them on the streets, as it were.

> But when I came inside, it was Gaelic folk who visited us, and I listened to them… But both have changed a lot. Everybody is distanced from each other, and they get to know each other through the devices… They're not close to each other at all. And it's the same thing in Glasgow as it is down in Kallin (Grimsay). People aren't – they're, as it were, isolated from each other, in Glasgow and in the Gàidhealtachd.' (Tormod MacGill-Eain)[5]

By no means a technophobe – he started his own blog in his eightieth year – Norman is not sparing in his contemporary critique. He keeps his critical eye open across the period of his lifespan, including on the young folk around him today in Grimsay where he ends up living.

5.4 Thursday: Creativity

Summary: Norman is invited to discuss his personal creativity as a teacher, writer, poet, musician, and comedian. He reflects on the varied influences of others, from backstreet singers to Billy Connolly, and discusses figures and trends in various art forms, and offers his opinions. He also recites a recently composed example of his own poetry, and other verses that have impressed him. In discussing how his bilingual background contributed to shaping his material, he also reflects on how commentators' propensity to place performers in pigeonholing categories could result in narrow or distorting descriptions of his work, for example as a 'Gaelic comedian.'

It is worth picking out the name Billy Connolly in this summary – firstly because when Norman first came across it, it was a barrier to him because of its Irishness. He was open about his prejudice, but he got over it and was quick to go on to say that this is the man from whom he gained the most influence as far as his comedic work was concerned. Once he had seen Connolly's act, that was the one he wanted to model his own on. He's also alive to stereotyping with respect to himself, through the use of social categorisations which he felt pinned into, while resisting at the same time. 'Gaelic comedian' would be the label that he would often be given even though he felt not only that there was no such thing, but in fact that there could be no such thing.

As a sample of his own creative output, the following extract gives a taste of his poetry.

> *Bheir mi luadh thar chàich*
> *Do bhailtean brèagha mo dhaoine:*

[5]https://youtu.be/NqUS1wwtCeA

Am Baile Sear is Bail' Mhic Phàil
Loch Euphort is An Caolas.
Innsidh mis' an-dràst',
Rim bheò cha dèan mi caochladh,
'S cha teirig dhaibh mo ghràdh,
Ged tha an teaghlach sgaoilte.

'I will praise above the rest
The handsome townships of my people:
Am Baile Sear and Bail' MhicPhàil
Loch Euphort and An Caolas.
I declare it now,
While I live, I will not change,
And my love for them won't fade,
Though the family is scattered.' (Tormod MacGill-Eain)[6]

This verse is selected from a poem which he recited during the recording, strongly demonstrating both rhyme and rhythm. I think he himself would say that his taste was more traditional than modern. He professed to having no great love for the work of Sorley Maclean, for example, the biggest upcoming name in Gaelic poetry at the time when he was at university. Thematically, the third and fourth lines are notable, being simply a repetition of a string of North Uist placenames. Perhaps there's some resonance there with the growing interest in the link between language and land, which seems to be attracting new scholarly and revitalisationist attention to some degree (e.g., Meighan & Chiblow 2021, Dziadowiec 2022).

5.5 Friday: Gaelic

Summary: On the last day, Norman is invited to turn his thoughts specifically to Gaelic and its place in people's hearts and minds, and to Gaelic development efforts. Acknowledging the challenges the language faces in today's world, he reflects on the complex interplay and relationships between Gaelic and English, and on various ways in which bilingualism can be viewed. In emphasising its benefits, he counsels against the dangers of a monolingual "English ghetto", colourfully invoking his own observations on the nomination campaign for the 2016 American presidential election. In contemplating bi-directional bilingualism, he discusses the challenges of, and offers his own advice on, the learning of Gaelic

[6]https://youtu.be/TlUzd4pKaIg

and, in particular, the place of literacy. Finally, he relates the language issue back to the culture from which it springs, sharing personal thoughts on how his sense of belonging reinforces his sense of identity, and emphasising his own willingness and commitment to pass on his knowledge to others.

The final Friday session, specifically on Gaelic, was another long one. I have chosen to highlight two quotes. The first illustrates his interest in bilingualism, an important subject of constant sociolinguistic debate and contestation (Jaspers 2016).

> *Thoradh, well, chan eil agad ach coimhead air mac Mhàiri Chaluim, 's e am bumalair sna Stàitean Aonaichte, Dòmhnall Trump… ach nuair a bhios tu a' coimhead air na daoine, aithnichidh tu… gu bheil iad, chan e dìreach gu bheil iad aineolach, ach gu bheil iad air leth, agus 's e is coireach – chan eil aca ach aon chànan. Tha sin gam fàgail, mar gum biodh, ann an* ghetto *Beurla… So, tha cunnart ann a bhith a' dèiligeadh le aon chànan a-mhàin. Tha e gad fhàgail – tha thu nas buailtiche a bhith fiadhaich agus a bhith air do dhalladh air an dealbh as fharsaing. Ma tha cànain eile agad tha an saoghal a' sìor leudachadh a-muigh, agus faodaidh tu roghainn a dhèanamh, faodaidh tu deagh thaghadh a dhèanamh.*
>
> 'Because, well, you only have to look at the son of Calum's Mary, that's the blockhead in the US, Donald Trump… but when you look at the people [his supporters] you'll see… not just that they're ignorant, but that they're isolated and the reason is – they only have one language. That leaves them, as it were, in an English ghetto… So, there's a danger in operating in only one language. It leaves you – you're more liable to anger and to be blind to the wider picture. If you have other languages, the world is always broadening out, and you can make choices, you can make good choices.' (Tormod MacGill-Eain)[7]

His substantive point here is that if you are restricted to one language your horizons are narrower than they need be. There is a danger that you get stuck in an "English ghetto." Stylistically, he makes the point more forceful by the reference to an important political figure. And he does it, first of all, through the *sloinneadh* rather than the man's English name itself. It is fairly common knowledge that Donald Trump's mother came from Stornoway in the Western Isles – so there is a Gaelic connection. It is that introduction of the individual through the *sloinneadh* which places him in a circle of assumed intimates, and

[7]https://youtu.be/IP2xWdottZM

so gives extra impact to the point. The special import that the *sloinneadh* carries says "We know who you are."

So he could be quite acerbic in his commentary, but at heart, there was a tremendous generosity there too, as the following final quote in this section demonstrates.

> So, *ma tha cuideigin a-muigh an sin, ma tha cuideigin a chì am film agad mu dheireadh thall, bhithinn gu math deònach na faclan agus na fuinn a chur thuige, thoradh fhuair mise an asgaidh iad, agus duine sam bith aig a bheil ùidh asta, gheibh iadsan an asgaidh iad cuideachd.*
>
> 'So, if there's anyone out there, if anyone sees your film eventually, I'd be very willing to pass the words and the tunes onto them, because I got them for nothing, and anyone who's interested in them, they'll get them for nothing too.' (Tormod MacGill-Eain)[8]

That was his sign-off. And it seems to me it is a good one, particularly given his role of tradition-bearer, and his felt need to pass this tradition on to the best of his ability.

6 Talking points with Norman Maclean

The Friday conversation saw the completion of *Saoghal Thormoid*. It is not quite the end of Norman's involvement in Island Voices, however, because another (posthumous) page, Talking Points, was added later. On that page, recordings of Norman have a central part.

We took short extracts from the Friday conversation, and we used them to spark sociolinguistic discussion between academic linguists in the University of the Highlands and Islands, the University of the West Indies, and Amity University Haryana in India, alongside UK-based Heritage Language speakers on the ground, of South Asian and Caribbean, as well as Hebridean backgrounds. So it is a bringing together of theoreticians, and of community practitioners whose professions include teacher, poet, and interpreter. There were about 20 minutes of *Saoghal Thormoid* extracts in total, which resulted in three and a half hours of discussion, all recorded and placed online.[9]

[8] https://youtu.be/IP2xWdottZM

[9] https://guthan.wordpress.com/talking-points/

We divided the discussions up into three broad topics – language endangerment, language hierarchies, and language contact – all of which could be of interest to language revitalisation practitioners and/or theorists. The following illustrative example is taken from the first discussion led by Ó Giollagáin, the lead author of the 2020 study on the "Gaelic Crisis in the Vernacular Community" (Ó Giollagáin et al. 2020). This substantial report is full of quantitative information of a much higher level of detail than is obtainable from census data. It is not a purely quantitative volume, however. Below are Ó Giollagáin's summary comments on Norman's seven-minute extract, consisting of the five main points which he took from it.[10]

- Implications of postwar societal change and modernisation
- Social densities of Gaelic speakers
- Minority youth socialisation processes in language contact
- Implications of the social subordination of the minority culture
- Differentiating language competence and vernacular communicative ease

All of these are important factors in his own sociolinguistic analysis, which he saw being corroborated in the qualitative ethnographic material presented in the *Saoghal Thormoid* recordings.

7 Other applications

7.1 Clilstore and DASG

Going on to other applications, what we have in *Saoghal Thormoid* (and other Island Voices collections) is good quality spoken material coupled with verbatim transcriptions. This can be used in a variety of different ways, including in the learner-friendly contexts which Clilstore was first designed to address (Ó Donnaíle 2013, 2022; Ó Donnaíle 2014). The online platform brings together media plus transcript plus dictionary look-up, available at a single click. You can watch and listen, you can read at the same time, you can click on any word that you do not know to get a translation.

This combination of audio plus transcript is useful in other contexts beyond language learning. So all the *Saoghal Thormoid* material transcribed for Clilstore

[10]https://youtu.be/krLY9ATKT5w

has gone into the Digital Archive of Scottish Gaelic (DASG) held at Glasgow University.

Given the orientation towards a vernacular model for use in new grammatical description (Ó Maolalaigh et al. 2014), this kind of material will be particularly useful, especially when it comes to developing a spoken corpus.

7.2 ASR and video subtitling

The project has partnered with the development of Gaelic Automatic Speech Recognition (ASR) at the University of Edinburgh (Evans et al. 2022). The benefit from this activity has been mutual. The ASR project needed good quality audio with transcriptions at the outset so that they could start building their language model and Speech To Text tool. Through its combination of well-recorded online video of Gaelic speech with verbatim written transcripts on Clilstore, *Saoghal Thormoid* (and Island Voices material more broadly) thus provided ideal material with which to begin that task. What Island Voices got back in return was a text alignment tool developed by the ASR project which helped us to start inserting subtitles into our YouTube videos automatically, albeit with a need for final manual adjustment.

So now we have Gaelic videos with optional Gaelic subtitling on Island Voices. As an added bonus, through Google Translate you are then able to secure automatic synchronous translation from Gaelic into scores of different languages as well.

8 Conclusions

This article has presented a sampling of one notable individual's varied contributions to a community-based exercise in open access online Gaelic language documentation. These contributions ranged across a variety of genres, including both scripted and unscripted speech. This provides valuable data in support of other Gaelic-related research and development projects, and a stimulus for cross-linguistic comparison with other Indigenous, minority or endangered language contexts.

Within this broad overview, special attention has been given to the subject's closing contribution to the project – the *Saoghal Thormoid* series of five conversations which Norman had with me over consecutive days in the late spring of 2016. Daily summaries have provided the gist of the overall content, while selected quotes and commentary have highlighted points of particular interest.

I have focused on three themes. I have sought to draw out Norman's observations on his own and others' use of Gaelic over the course of his life. At the same time these observations have been placed alongside his wider running commentary on the changing social context as he has observed and understood it from childhood through to old age. And in particular, I have highlighted the significance of the Gaelic *sloinneadh* as a key traditional marker of identity within the community context, and how Norman has used it recurrently in the conversations.[11]

Island Voices, as an ongoing project, remains open to further collaboration and partnership opportunities in other related fields. There's certainly plenty more to investigate, for example around the topic of bilingualism. The video used as a final sign-off clip in the Talking Points series shows an example of real-time code-mixing between Gaelic and English from Norman, who was perfectly fluent in both languages.

The YouTube description underlines this linguistic complexity, while also paying tribute to the cultural wealth of his contributions:

> "… Notably, from a linguistic point of view, the clip features frequent code mixing and/or switching. The subtitled translations from Gaelic into English are in plain script, while English sections (or single words) are in italics. (The occasional use of Gaelic place or personal names in an English sentence is not distinctively marked.)
>
> Aside from these points of linguistic interest, we can count ourselves very lucky to have been able to record Norman for posterity, not only - as here - soon after he returned to Uist, but again and again after he settled. He was extremely generous with his time, and became a prolific contributor to Island Voices, particularly towards the end of his life, offering copious comic relief as well as serious insight into his life and times." (Island Voices Videos)[12]

[11] I have suggested that my own *sloinneadh* may be construed as part of my positionality within the community and within this project. To further suggest that this community knowledge, familiar to Norman, may have contributed to his ease in speaking openly and naturally about topics of interest to him might be presumptuous. Venkatesh (2002), and Jaspers & Meeuwis (2013) are admirably self-aware and self-critical in discussing issues around the researcher's own profile and investment in community-based enquiry. Space has not allowed extended discussion of related issues in this piece, but further research and indeed introspection on this topic on the part of Gaelic scholars with an interest in the social contexts of its contemporary use may well be warranted.

[12] https://youtu.be/XFGInaf_DGQ

Norman's talent was remarkable, as was his generosity. He gave us plenty to think about, but he allowed us to enjoy thinking about it at the same time through his lively presentation skills. He was an exceptional exponent of Gaelic, as well as linguistic competence more generally. This brings me back to the conundrum posed in the introduction. What lessons can we take from exceptional exponents when pursuing the study of the language behaviour of ordinary people?

My answer to this question in respect of Norman Maclean, specifically, is in two parts. First, he was a brilliant comedian, and I do not think you can make people laugh the way he did if you do not speak to them in some sense "in their own language" – if you do not reach them and make an affective connection. And he certainly did that. Alongside his many talents he was firmly grounded and empathetic. So his performative voice was at the same time a demotic one – being 'of the people.' In this respect, Norman's Gaelic speech patterns may well be taken as fairly representative of the kind of "retro-vernacular" model preferentially selected, for example, in corpus-related linguistic study (Ó Maolalaigh et al. 2014).

The second point is a methodological one. The recording format that we developed in collaboration with Norman for *Saoghal Thormoid* (Wells 2017) is one which we have since adopted for wider work. The *Stòras Beò* 'Live Store' page tab on Island Voices to a large extent replicates the *Stòras Beò nan Gàidheal* 'Live Store of the Gaels' series being collected on the webpages of the University of the Highlands Language Sciences Institute.[13] In this collection, we've used that same recording model, putting separate cameras on two people, and letting them talk on topics of interest. The resultant recordings are edited together with a light touch, and then placed online for open access. So, in terms of the format, Norman's input to *Saoghal Thormoid* may be regarded as a pioneering blueprint for later dissemination, preparing the ground for many more community members to take the plunge and participate as interviewees, interviewers, or both.

Whether wittingly or not, through his infectiously enthusiastic engagement and prolific output, Norman enhanced and advanced the work of Island Voices immeasurably, for which we can only be forever grateful.

References

Bloomfield, Leonard. 1933. *Language*. New York: Holt.

Casey, Louise. 2016. *The Casey review into opportunity and integration*. Tech. rep. Department for Communities & Local Government, UK Government. https://www.gov.uk/government/publications/the-casey-review-a-review-into-opportunity-and-integration (5 June, 2025).

[13] https://www.uhi.ac.uk/en/research-enterprise/res-themes/humanities-and-arts/language-sciences-institute/projects/storas-beo-nan-gaidheal/

Chafe, Wallace & Deborah Tannen. 1987. The relation between written and spoken language. *Annual Review of Anthropology* 16. 383–407. DOI: 10.1146/annurev.an.16.100187.002123.

de Saussure, Ferdinand. 1966. *Course in general linguistics*. Translated by W. Baskin. New York: McGraw-Hill.

Dziadowiec, Ryan. 2022. *Ceist an fhearainn, ceist a' chànain, agus an anthropocene: Freagairtean à lèirsinn ghàidhealach (the land question, the language question and the anthropocene: Solutions from a Gaelic perspective)*. Soillse. http://www.soillse.ac.uk/wp-content/uploads/Dziadowiec-Soillse.pdf (5 June, 2025).

Evans, Lucy, William Lamb, Mark Sinclair & Beatrice Alex. 2022. Developing automatic speech recognition for Scottish Gaelic. In *Proceedings of the 4th Celtic language technology workshop within LREC2022*, 110–120. Marseille: European Language Resources Association.

Greenbank, Paul. 2003. The role of values in educational research: The case for reflexivity. *British Educational Research Journal* 29(6). 791–801. DOI: 10.1080/0141192032000137303.

Jaspers, Jürgen. 2016. Diglossia and beyond. In Ofelia Garcia, Nelson Flores & Massimiliano Spotti (eds.), *The Oxford handbook of language and society*, 179–196. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780190212896.013.27.

Jaspers, Jürgen & Michael Meeuwis. 2013. Away with linguists! Normativity, inequality and metascientific reflexivity in sociolinguistic fieldwork. *Multilingua* 32(6). 725–749. DOI: 10.1515/multi-2013-0035.

Meighan, Paul & Susan Chiblow. 2021. Language is land, land is language: The importance of indigenous languages. *Human Geography* 15(2). DOI: 10.1177/19427786211022899.

Ó Donnaíle, Caoimhín P. 2013. Multimedia-assisted content and language integrated learning. *Multimedia-Assisted Language Learning* 16(4). 11–38. DOI: 10.15702/mall.2013.16.4.11.

Ó Donnaíle, Caoimhín P. 2014. Tools facilitating better use of online dictionaries: Technical aspects of multidict, wordlink and clilstore. In *Proceedings of the first celtic language technology workshop*.

Ó Donnaíle, Caoimhín P. 2022. CLILSTORE.EU: A multilingual online CLIL platform. In *Proceedings of the 4th Celtic language technology workshop within LREC2022*, 22–29. Marseille: European Language Resources Association.

Ó Giollagáin, Conchúr, Gòrdan Camshron, Pàdruig Moireach, Iain Caimbeul, Brian MacDonald & Tamás Péterváry. 2020. *The Gaelic crisis in the vernacular community: A comprehensive sociolinguistic survey of Scottish Gaelic*. Aberdeen: Aberdeen University Press. DOI: 10.57132/mpub.14497417.

Ó Maolalaigh, Roibeard, Mark McConville, Wilson McLeod & Susan Bell. 2014. *Dlùth is inneach – final project report: linguistic and institutional foundations for Gaelic corpus planning*. Tech. rep. Bòrd na Gàidhlig. https://www.gaidhlig.scot/wp-content/uploads/2016/12/%C3%93-Maolalaigh-et-al.-2014-Planadh-corpais-CR12-03-Corpus-planning.pdf (5 June, 2025).

Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt-Brace.

Scotland's Census. 2011. *2011 census: Key results from releases 2a to 2d*. https://www.scotlandscensus.gov.uk/documents/2011-census-key-results-from-releases-2a-to-2d/ (5 June, 2025).

Venkatesh, Sudhir. 2002. 'doin' the hustle': Constructing the ethnographer in the American ghetto. *Ethnography* 3(1). 91–111. DOI: /10.138102003001004.

Wells, Gordon. 2017. *Saoghal Thormoid – Norman's world: Synopses and transcripts*. Sleat: Soillse. https://guthan.wordpress.com/research/ (5 June, 2025).

Wells, Gordon. 2020. Reading Island voices: Issues around the primacy of speech and the privileging of literacy, from a Hebridean viewpoint. *Changing English: Studies in Culture and Education* 27(1). 109–117. DOI: 10.1080/1358684X.2019.1660620.

Wells, Gordon. 2023. *Island voices – guthan nan eilean: Hebridean language capture and curation, 2005-2023*. Benbecula: Island Voices. https://guthan.wordpress.com/research/ (5 June, 2025).

Widdowson, Henry. 1998. Context, community, and authentic language. *TESOL Quarterly* 32(4). 705–716. DOI: 10.2307/3588001.

Chapter 14

White supremacists' weaponization of Celtic heritage languages and language revitalization

John Powell
University of Arizona

In response to the rise in white supremacist terrorism and hate crimes, there has been an imperative to monitor white supremacist networks. Significant linguistic scholarship has examined how white supremacists espouse linguicidal ideologies of English monolingualism and oppress users of minoritized dialects and languages. Linguists and language activists have combated this white supremacy and colonialism by promoting heritage languages and language revitalization, but little work has examined how some white supremacists appropriate and weaponize these efforts, especially with Celtic languages. This paper examines language ideologies in corpora of white supremacist speech. The findings show that some white supremacists support and learn their languages of heritage to index their white supremacy, which include minority and even endangered languages of Europe, especially those in the Celtic family.

1 Introduction

Before I proceed with my introduction, it is prudent to provide both a warning of the content and a note on my position as a scholar investigating this topic.

This paper contains uncensored quotes (with references to source material) from corpora of hate speech extracted from white supremacism, white supremacist and extremist sites, and some of these quoted examples include slurs.[1] I leave

[1]There are no quotes that include the n-word in this paper.

John Powell. 2025. White supremacists' weaponization of Celtic heritage languages and language revitalization. In Andrew Carnie, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond & Luis Irizarry (eds.), *Foundational approaches to Celtic linguistics*, 495–525. Berlin: Language Science Press. DOI: 10.5281/zenodo.15654883

them fully quoted so they can be viewed in context. I also do not fix any typos or grammatical errors in the quotes.

I recognize that ideologies of language quoted in this paper have traumatized countless people and communities. While I acknowledge that some may be (re)-traumatized by this hate speech, this is not my intention, but rather to expose, analyze, and fight white supremacists' ideologies on language.

Regarding my position, I am a straight white cis male of Scottish, Irish, and Welsh ancestry, whose first language is a privileged variety of American English and whose second language is Brazilian Portuguese, acquired as an adult.

As a linguist, I am a proponent of learning heritage languages and language revitalization. Indeed, some of these are major areas of research for me, including my own heritage Celtic languages. However, this paper will show how these areas are being appropriated and weaponized by white supremacists for their own purposes.

1.1 Background

There has been a rise in white supremacist terrorism and hate crimes (Byman 2022). As a consequence, there has been an imperative to monitor white supremacist networks. Scholars in various disciplines have recently embraced this imperative, especially in light of the observation that white supremacist organizations have been "zealous consumers and promoters of primary scientific literature" (Carlson et al. 2022: 445) for racist ends. Recent scholarship in different disciplines has sought to expose and counter the presence of white supremacist subcultures within their broader fields, such as in folklore (Wilson 2021), ecology (Bergman 2021), genetics (Carlson et al. 2022), among many others.

Throughout the last several decades, linguistic scholarship has examined how white supremacists espouse linguicidal ideologies of English monolingualism and oppress users of minoritized dialects and languages (Kroskrity 2021, *inter alia*). Linguists and language activists have combated this white supremacy and colonialism by promoting heritage languages and language revitalization (Hinton et al. 2018). However, very little work has examined how some white supremacists appropriate and weaponize these efforts, which is the subject of the present chapter.

There is robust literature of looking at hate speech on white supremacist websites. Daniels (1997: 37) found that:

> On almost every front in which white men are presented as "warriors" they are simultaneously presented as "victims." White men depict themselves

> as victims of racial discrimination, of class oppression, and as the special victims of race, gender, and class oppression at the hands of the racial state.

Bridle (2016) conducted a critical discourse analysis on posts on Stormfront, a white supremacist forum. He focuses particularly on homophobia. He finds that despite the fact that the:

> overwhelming majority of discourses and texts on Stormfront related to gay men are both negative and homophobic, […] race takes precedence over sexual orientation, as gay men are constructed positively on the condition that they support white nationalism (p. 92).

Brindle concludes that, for example, white supremacists rank gay men above Muslims.

Yet, there has been little research that has looked at white supremacist websites as primary sources within the fields of heritage language, language endangerment, and language revitalization, other than white supremacists opposing and oppressing these efforts.

1.2 Framework

I am utilizing part of Rosa & Flores's (2017) framework of raciolinguistics, which frames "the co-naturalization of language and race as a process of raciolinguistic enregisterment, whereby linguistic and racial forms are jointly constructed as sets and rendered mutually recognizable as named languages/varieties and racial categories" (p. 631). Critical to their framework are the "continued rearticulations of colonial distinctions between Europeanness and non-Europeanness – and, by extension, whiteness and nonwhiteness" (p. 622). This is similar to how Irvine & Gal (2000: 73) discuss the "dichotomizing discourses of orientalism" which is seen in ideologies on languages, and where Europe is viewed "in opposition to a broadly defined 'East'". We will see that these are the sorts of distinctions that white supremacists seek to reinforce in what languages they find acceptable to be spoken, and even more, who is allowed to speak them.

I am also employing intersectionality (Crenshaw 1989, inter alia). Throughout this paper, we see how white supremacists punish every departure from the straight white cis male "native" speaker "standard," whose speech is the only kind white supremacists find acceptable.

1.2.1 Research questions

My research questions are as follows:

(a) What languages do white supremacists view as acceptable? Which ones do they view as "for whites only"?

(b) What are their ideologies toward language diversity, minority languages, heritage languages, language endangerment, and their revitalization?

(c) How do they view Celtic languages?

(d) What are the consequences of their ideologies?

This paper examines language ideologies in corpora of white supremacist and extremist speech.

1.3 Data source

I used three primary sources for my corpus analysis:

- de Gilbert et al. 2018's *Hate Speech Dataset* (HSD)
- Kennedy et al. 2022's *Gab Hate Corpus* (GHC)
- Metapedia[2]

I will address each primary source individually.

1.3.1 Hate Speech Dataset (HSD)

The HSD is a corpus of extracted text from a sample of 5,000 Stormfront posts (around 0.05% of the site), which is a major white supremacist forum and the oldest and most influential hate website (Hartzell 2020).[3] HSD contains 196,571 words across 21 subforums collected between 2002 and 2017.

I am looking at this hate-filled corpus because it provides explicit insight into white supremacists' ideologies. It implicates all posters as white supremacists or white nationalists, whereas other message boards with numerous racists (i.e.,

[2] www.en.metapedia.org

[3] Southern Poverty Law Center (2022) considers the Daily Stormer more influential recently, but to my knowledge, the one study which examines posts on the Daily Stormer (Calderón et al. 2021) has not made their corpus public.

4chan, 8chan/8kun, etc.) aren't *exclusively* defined by white supremacy. No one is going to post on a website whose banner reads "white pride world wide" unless they espouse white nationalist ideologies. Finally, according to Beirich (2014), registered users have been connected to over one hundred murders.

One important methodological note is that HSD samples posts throughout the subforums and that the samples comprise only single posts, not dialogue or conversations. Thus, the context of a given post is not always immediately clear. However, metadata in HSD provides information of which subforum each post comes from, so general themes and topics surrounding a given post can be ascertained.

It is also important to note that while Stormfront is hosted in the United States, and American white nationalists are a massive group on the website, it is very much an international venue for hate speech. In fact, the slogan on its banner shows this was always its mission. As such, the forum has many posters from all over the world (though no metadata on location exists in the corpus).

1.3.2 Gab Hate Corpus (GHC)

Gab is a social media platform with a far-right user base – a Twitter of sorts for extremists. The GHC is a corpus of extracted text from 27,551 posts (a small fraction of the millions of posts) from Gab, collected between January and October of 2018. The corpus contains 564,280 words. Like the HSD, the corpus does not preserve dialogue or conversations, but rather individual posts.

I am looking at a corpus of Gab because the platform has become a space for far-right extremists, notably white supremacists, conspiracy theorists, right-wing insurrectionist and militant types, and fascists. Because Gab explicitly refuses to moderate its platform, many extremists who were deplatformed or otherwise disaffected from other social media websites have migrated to Gab, which has become a "haven for dangerous white supremacists" according to the editors at the Southern Poverty Law Center (2019). David Duke is a major figure on this website, and it was the preferred social media platform of domestic terrorist Robert Bowers. This social media site provides real-time reactions of extremists to current events.

Nevertheless, unlike the other two data sources, Gab isn't exclusively defined by a white supremacist userbase, but it is the case that white nationalists are a principal group, and perhaps the largest. So, while the userbase of Gab isn't categorically white supremacist, unlike Stormfront, it is certainly stochastically white supremacist. A simple search of the term "Jew," "white genocide," Nazi slogans, and the n-word in GHC reveals the ubiquity of white supremacist ideology

proliferating on the site. Further, posts from GHC analyzed in this paper were chosen for their specific espousal of white nationalist and racist ideology, which is different from HSD, where all posters are implicated as white supremacists by simply being users of Stormfront. Otherwise, I characterize the posters on GHC as extremists, which is an accurate characterization of its user base.

1.3.3 Metapedia

Metapedia is a white nationalist wiki founded in Sweden. According to SPLC, Metapedia bills itself as "an alternative encyclopedia dedicated to the pro-European cultural struggle." Metapedia has a number of articles on various languages, as well as articles on linguistics and languages. I use computational tools to scrape articles from the website.

I am examining articles on Metapedia because, as a wiki, the website provides a window into white nationalist ideologies and philosophies on a number of subjects. Metapedia has been examined in other fields for windows into racist pseudoscience (de Keulenaar et al. 2022). According to Oeberst et al. (2018), Metapedia articles tend to be short and unsourced, and de Keulenaar et al. (2019) further characterize the articles as "alternative facts."

It is important to note that many of the articles on Metapedia are stubs and fragments, and the distribution of topics is uneven and varies greatly by language. Arnstad (2015) found the distribution of articles by languages on Metapedia seen in Table 1. All articles were scraped in November 2022.

1.3.4 Other sources

I also consulted two major databases of white supremacist terminology, symbology, and ideologies. The first is the Anti-Defamation League's 2022 *Glossary of Extremism* and *Hate Symbol Database*, which is a source that includes symbols, images, and descriptions on 214 hate symbols. The second is the Southern Poverty Law Center's 2022 *Extremist Files*, which includes databases on 136 extremists, 72 extremist groups, and 24 ideologies.

2 Methodology and results

In this section, I explain each of the computational methodologies I employ for each of the three primary data source corpora.

Table 1: Number of articles by language on Metapedia

Language	No. of Articles
Hungarian	146,470
German	55,859
English	19,862
Swedish	10,271
Spanish	9,167
Estonian	2,692
Romanian	2,393
French	1,878
Portuguese	1,629
Slovenian	935
Norwegian	602
Dutch	412

2.1 Hate Speech Dataset

I looked at what languages white supremacists talk about. So, I used computational tools to find language keywords from posts on HSD. Using a comma separated value file of 12,697 names of dialects, languages (both current and historical), and language families pulled from lists on (the real) Wikipedia, I wrote python code that, with regular expressions (*re*) patterns `(^X.*)`, searched names for languages in the corpus, including derived and inflected forms.

Note that Irish was the second most frequent language mentioned, and Celtic the most frequent mentioned family.

Next, I looked at what languages white supremacists use in their posts. It seems that moderators on Stormfront routinely police subforums and delete non-English posts, as seen in the following quote.

> Gentlemen, I remember you that all the posts not in English will be deleted. Only posts in English, please!??? [...] (HSD: 14076260)[4]

As expected, the posts are overwhelmingly in English. But moderators occasionally tolerate posts in other languages. So, I identified the language of every post

[4]Every quote is provided with a numerical reference of the source corpora on HSD or GHC in parentheses, or in the case of Metapedia, the name of the article.

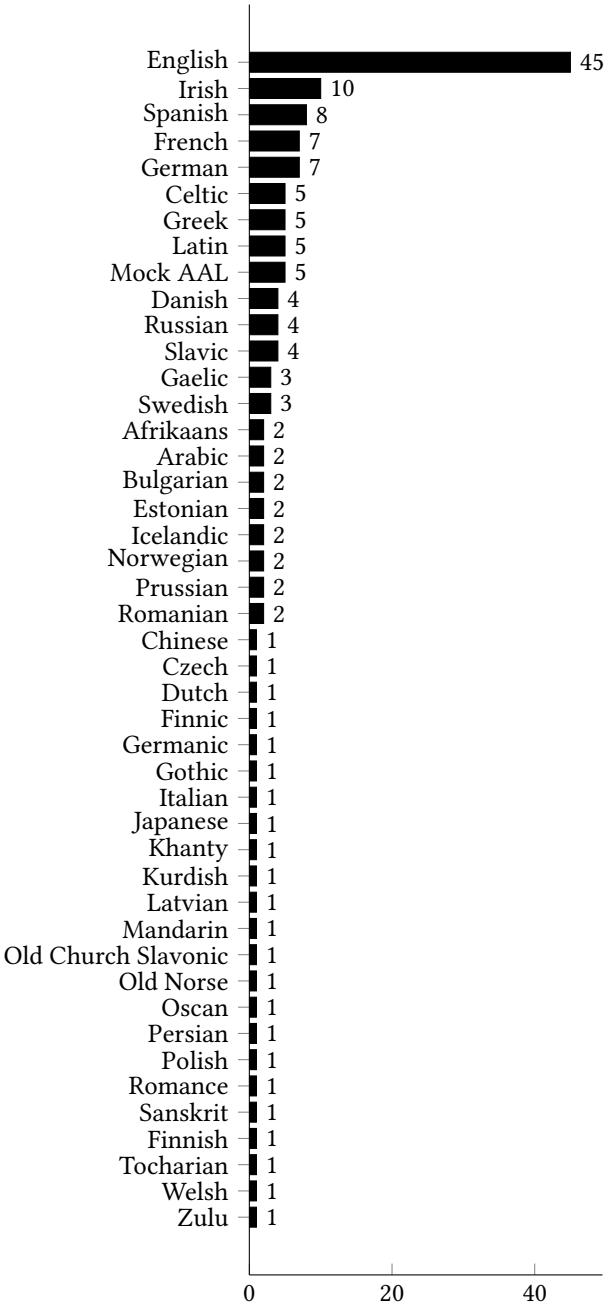


Figure 1: Frequency of the mentions of languages in HSD

using `langid.py` package (Lui & Baldwin 2012).[5] This provides a scope of languages tolerated on Stormfront.

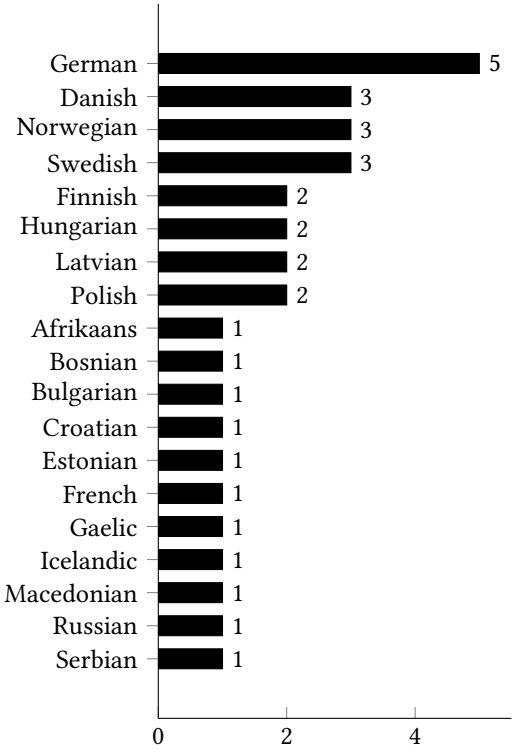


Figure 2: Frequency of posts in languages besides English on HSD

All non-English posts came from European languages (exclusively from the Germanic, Uralic, Slavic, Celtic, and Romance families). As such, Stormfront enforces English-only policies but that the universe of tolerated languages are all European or descendants from European languages. In Section 3, I return to the contents of the posts.

2.2 Gab Hate Corpus

I used the same code for examining GHC. However, as I mentioned, Gab is defined by its extremist user base, and as such, since this is a study about and not extremism more broadly, I use this code for GHC for different purposes.

[5] I removed posts that included rote Nazi slogans in German, as well as advertisements.

First, I looked at what languages are generally used on Gab. An important distinction between HSD and GHC is that the latter doesn't purge non-English posts.[6] As a result, there are subcommunities on Gab which would be characterized as far-right groups in a number of other countries, especially those in Germany and Brazil associated with Alternative für Deutschland (AfD) and Partido Liberal (PL), respectively. Running `langid.py` revealed 174 non-English posts, about 0.63% of the entire corpus. Of those posts, 75 are in German, 59 in Portuguese, 14 in Dutch, 13 in Spanish, 9 in French, 2 in Latin, 1 in Bosnian and 1 in Romanian.

Then, I examined what Gab users, especially those who explicitly espouse white supremacist ideologies in their posts, say about languages, especially Celtic ones. I examine these along with posts from HSD in Section 3.

2.3 Metapedia

Finally, I wanted to find out what sorts of topics related to language and languages, especially those discussed in this study, are found within Metapedia articles. So, I searched for and scraped articles on Metapedia. Metapedia uses the same general API that all Wikis use. Thus, I retooled code for scraping Wikipedia articles to scrape articles on Metapedia. Specifically, I used the *requests* library for python for the Metapedia API. I performed search queries (using the *query* function) of key terms in article names, usually involving "language," "Celtic," or other related topics. Most of the articles on language feature the keyword language in the article title, and many of the title names reflect those of articles on Wikipedia, so that helped me in crafting optimal searches. During this process, I would also note hyperlinks and related articles to expand my search. Then, I scraped (using the *parse* function) HTML of articles on the website into a *json* format text output. This was conducted all behind a commercial VPN. I did this in order to avoid visiting Metapedia in my browser, so I am thus merely working with text outputted from my python code. As such, I wasn't interacting with Metapedia articles the way typical users do, but I was nevertheless able to browse and extract text from the site.

I qualitatively explore the texts retrieved from these computational tools in the subsequent sections.

[6]However, many users exhibit hostility towards those who don't post in English on Gab.

> Fucking foreigner. I don't deal with people that can't speak proper English. (GHC: 20650)
>
> if you're trying to make people stop using your app, continue to shove all these foreign language posts in our faces. I am getting tired of scrolling over incomprehensible foreign shit. (GHC: 12480)

3 Discussion

Before I discuss posts related to Celtic languages, I must first dispel the notion that because Stormfront and Gab are predominately associated with American white supremacism and extremism, that they only promote English monolingualism. The situation is far more complicated.

3.1 English-only ideologies and whiteness

To be sure, plenty of posts on HSD promote English-only ideologies, and even English supremacy.

> […] In the battle of languages English has won. (HSD: 13440452)

White supremacists discourse often involves appeals to conquest of white people over other races, and this post mirrors that for languages. Yet, posters routinely complain about their perception of how many people, largely immigrants, don't speak English.

> its the same way in denver. i dont even think anybody speaks english out here anymore. i love this town but i think its time to go... (HSD: 14109242)[7]

As Stormfront is an international venue for white supremacism, the English-only ideologies are not always expressed from an American perspective, especially evident in the following quotes.

> Yeah iam from Edinburgh and i when i was at secondary school my classes where full of blacks and pakis and half them could hardly speak english. (HSD: 14030845)
>
> A British school where none of the students have English as their first language? This is the last straw, how much more will damage Europe and Western civilization these Zionists? (HSD: 14034403)

The previous quote makes explicit that anti-immigration ideologies of white supremacists are rooted in antisemitism. This is evident in their orientalist discourses towards a school with second language speakers. Moreover, there is a reference to the great replacement conspiracy theory of Jews damaging "Europe"

[7] As noted above, I do not correct spelling or punctuation errors in quotes from postings, nor do I censor offensive language. I present them here verbatim.

and "Western Civilization" through the education of immigrants, especially those from outside Europe and the West.

Some posters put forward notions that immigrant students receive more attention than non-immigrants, without any evidence.

> Irish pupils are getting less teacher time too because they're so busy helping the immigrant chrildren get up to date. (HSD: 13858717)

Their solutions often include repeating oppressive histories, as seen in the following posts.

> I agree. And that goes for Eastern Europeans and Asians as well. All schools in Ireland should be segregated. If teachers have to focus extra attention on foreigners because they cannot speak English properly, it will be the indigenous Irish students who will suffer. (HSD: 13855938)
>
> they sent many of the irish immigrants over hundreds of years ago during one of the first ever drafts I beleive. So lets do it to the beaners. (HSD: 14414751)

White supremacists on HSD cite these oppressive histories toward the Irish to not only diminish the plights of other immigrants, but also to justify or even repeat historical trauma onto them, including treatment that the Irish historically suffered, as seen in the mention of segregation (in the Irish sense) or forced relocation.[8]

The "replacement" ideology extends explicitly to language.

> I concur with you totally, it's a sick county I live in. They said that statistics show that by the year 2011 Espanol will be the main spoken laungage. (HSD: 31688744)

This post, likely created sometime between 2002 and 2010, is suggesting that English will be replaced by Spanish in some county, despite the fact that the opposite occurs – more bilingual Spanish communities are shifting to English monolingualism in the United States. Yet the ideology behind the post hearkens

[8] One poster even muses about expanding racial segregation in classrooms by reversing the tools of integration seen in minority-majority schools.

> I want to see Whites bused in there until the Latinos are a 30% minority and not outnumbering the Asians, Blacks and Whites. Where is forced busing when you need it? (HSD: 13209079)

again to the great replacement theory that the diminishment of the language is evidence of the replacement of white people. We see the term "replacement" used explicitly in a post on GHC, where English is viewed not only as a minority language, but incredibly one that needs special protection like Quebecois French.

> "Get with the program, racists, and learn Spanish." The "program"-displacement of Historic American Nation. Quebec-type language laws the only answer (GHC: 23184)

Some posters do not consider speaking English enough, but that immigration of any kind results in the so-called "replacement" of white people, as seen in the post on GHC.

> With respect ur 100% wrong we need to be worried about ppl abusing the right wing anti immigration sentiment and moving it into the direction of support for soft multiculturalism. "mass immigration is ok as long as the immigrants speak english and dont rape kids" sentiment tjats wot Tommy stands for its sick. Replacing our ppl is nva ok (GHC: 6973)

White supremacists view a speaker's race as deterministic in their ability to speak English, as seen in the post below which rattles off a number of activities which the poster associates with white people, and don't associate with people of color, including speaking English.

> I never really thought about it before but I guess bicycling is *primarily a white activity*, kind of like sunbathing, going swimming, playing golf, tennis, attending a social function without a brawl taking place, not smelling like a wild animal in the locker room after a hard workout, and *speaking proper English*. (HSD: 30444129, *emphasis mine*)

In each of these cases, white supremacists reinforce racial barriers in their determination of who can speak English.

3.2 European languages and whiteness

From the quotes so far, it might be easy to assume that white supremacists only promote English monolingualism and English-only ideologies, at the expense of any other language. But that is often not the case. Many on HSD indeed promote learning additional languages.

> Raise your kids on more than one language. Apparently young kids pick languages up very quickly. I speak two. So do my kids, and they are 4 & 5. (HSD: 30747632)
>
> It is a good language to learn and is related to all other Slavic languages. […] (HSD: 13504387)

Other posters on GHC echo these sentiments, which is indicative of the embrace of languages among some far-right extremists.

> Read old books. Learn new languages. Good advice. Good article. (GHC: 25192)
>
> If you want to learn a new language, you don't have to move to a foreign country to immerse yourself in the language. You can stay immersed just by reading and listening to it all day no matter where you live. (GHC: 16755)

Often, white supremacists discuss learning their own heritage languages, or those which, for the purposes of this paper, I define as the languages which individuals have ancestral connection to, regardless of whether they grew up hearing the language or not. Such seems to be the case for this poster on Stormfront.

> Slowly making my way through a Romanian Bible I got for my birthday in July. Though I am not a Christian, I can read it alongside the English and try to figure out words I don't know. (HSD: 30630091)

Nevertheless, for white supremacists, the only languages worthy of learning are European languages. It is clear from posts in HSD that the white supremacists' motivation for learning languages is often related to their white heritage. We can see this in several examples. First, we see it with a user on Stormfront who was crowdsourcing Danish census terms in order to conduct research on family genealogy.

> I would like to finish this tree for posterity, so what about these terms? Kokkepige Stuepige Barnepige Tjenestepige Tienestefolk Huus jomfure (House something?) Politifuldm. Står i butik (papirforretn.) (The above are other people listed as living in my family's home. Servents/maids, workers of some type? Lodgers?) Also: Student, underholdes af faderen (Student of ... ?) Handelsbetjent I really appreciate your help, Profoss. It took me forever to find these documents.... (HSD: 13590117)

Another engages in a discussion about Finno-Ugric (Uralic) folklore and songs.

> Please, enough of the Finno-Ugric fairy tale […] And here is a Khanty song for comparison Khanty song "Akem ar" ("Old Man's Song") (HSD: 12872207)

And several others trace etymologies of words in Indo-European languages.[9]

> Yup. Here's a list of some of the other derivitives: Skr. धूम (dhūma), Old Prussian dumis, Russ. дым (dym), Lat. fūmus, Gm. toup/toben, Gk. θυμός (thumos), Toch. twe/tweye, Lith. dūmai, Ltv. dūmi, Eng. dōmian/dusk, Goth. dauns, ON daunn, Ir. dé/deathach, Welsh dywy, Oscan Mefit, Polish dym, Pers. dud, OCS дымъ (dymŭ) (HSD: 30586603)

While the content of this post appears benign, this is nevertheless one poster on Stormfront accessing cognate sets in Indo-European languages. It shows that the tools used by linguists are appropriated by white nationalists, though to what end isn't clear in this post without more context.

Not every white person using their heritage language is praised.[10] One post expressed disdain for those who use their language but don't promote white supremacy.

> They [the Irish] are a good example of people who are brainwashed. They speak the Irish language because it's their culture but are brainwashed into thinking that it's ok to destroy our race. (HSD: 13863902)[11]

> Ireland is run by England/Jews. The Irish make no law or policy. Nor does Sinn Fein Or Ira. (GHC: 2999)

This is one of the few instances where white speakers are criticized, but it is relatively uncommon. Most of white supremacists' scorn is reserved for the speech of people of color.

And just like how white supremacists only accept white people speaking English, many only accept white people speaking a language originally from Europe.

[9] Other examples include *ingratiate* (HSD: 30734623), *epistrophe* (HSD: 30791003), *lambent* (HSD: 30704903), and Old English and Celtic origins in the surname *Lee* (HSD: 30620846).

[10] Indeed, one post from 2018 on GHC expresses disdain for the protection of Irish in Ulster in what appears to be Unionist support of English monolingualism.

> PLEASE SIGN THIS PETITION ON BEHALF OF GOD'S PEOPLE IN BRITISH ULSTER NOW UNDER THE THREAT OF A LEPRECHAUN LANGUAGE ACT!!!!!! DON'T LET THE VERMIN GET AWAY WITH CULTURAL GENOCIDE!!!!!! (GHC 3382)

[11] We see similar kinds of bemoanings rooted in antisemitism in the GHC.

> Spanish is NOT their language! It is our (White) language, just like French is our language even though millions of negroes in Africa speak it. (HSD: 30756516)
>
> There's something so gross and such an assault on the senses as the sound of an European romance language coming from the mouths of savages! (HSD: 13384830)

In the post above, white supremacists attribute ownership of both Spanish and French to white people, even though the majority of the speakers of both languages are people of color. And these are the speakers of Romance languages that the white supremacists show disgust for. The latter post does not mention how people of color speak the language, or their varieties or dialects, but rather the posters' animosity seems directed squarely at their race – that people of color have no right to speak the language or aren't even able to speak the language. These posts show that it's not even about language per se, but rather the race of the speaker, regardless of language. White supremacists always punish non-white speakers, whether they speak a European language or a non-European language, and this falls out from intersectionality.[12]

In posts discussing national languages of various countries, this notion of who can and can't speak the language is reinforced, couched in an ideology of linguistic nationalism, not unlike the English-only posts.

> […] For most of the service jobs you definatly have to speak Danish well. So that excludes uneducated third worlders. (HSD: 13591809)

Posters on Stormfront from various countries frequently and forcefully demand that immigrants leave, making explicit their racist appeals to their national language, motivated by white nationalism.

[12]White supremacists even folk theorize that the supposed mutual intelligibility of dialects is determined by the race of the speakers.

> Americans and Irish understand each other better than Spaniards and the beaners … the US and Ireland are the 1st world so they are educated unlike the beaners. (HSD: 30691153)

The poster advances the notion that somehow mutual intelligibility is related to education, but the use of the slur for Hispanics betrays the poster's ideology – that they view race as deterministic for mutual intelligibility. This is rooted in ideologies of dialect and language supremacy, that is multiracial speakers of languages are less likely to understand the "pure" white varieties and vice versa.

> everyone who doesnt speak Romanian and look as in WHITE get out of Romania a move back to your homeland. Russians. Hungarians. Gyspsies! Jews! Dessert ******!! (HSD: 14077746)

This disturbing post shows how white supremacists view race, language, and nationalism as key components of their racist ideology, where, to this one poster for example, the only acceptable people living in Romania are those who were born in the country, speak the language, are not Jewish or Roma, *and* are white.

This racist ideology parallels how white supremacist view race and nationality, where posters routinely exclude people of color from definitions of who is European and what is their nationality.

> In certain circumstances I would agree with you but in Europe by preserving our nations we are preserving our race as well because there is no way a black in Europe can be British, Irish, German or whatever. (HSD: 13859163)

This reinforces their racist distinction of Europeanness and non-Europeanness (Rosa & Flores 2017), where the criteria is some unarticulated notion of ancestorial heritage which gatekeeps those who can and can't speak a language. Not only is this a co-naturalization of race and language, within their theory, but seemingly one where white supremacists view the capacity to speak a language as one which is inherited genetically.

Just like in the English-only context, white supremacists often view other languages in their country as some twisted form of evidence of their 'replacement.' Metapedia articles consistently reiterate this ideology. In the article on the Arabic language, it states:

> [...] Arabic is the language of the Quran and the religious language of all Muslims, therefore having been spread in association with Islamization of various areas. (Metapedia: *Arabic language*)

The authors of the article first conflate Classical Arabic of the Quran with Modern Arabic. Notwithstanding, the article then bizarrely associates the use of the language as evidence of Islamization, which is frequently found in their replacement discourses.

For white nationalists, it is always the linguistic replacement of European languages, ever the victims. We see the same kind of perspective in another Metapedia article, but this time with a German dialect.

> […] in the French regions of Alsace and Lorraine French has *replaced* the local German dialects as the official language, even though it has *not been fully replaced* on the street. […] Some German-speaking communities still survive in parts of Romania, the Czech Republic, Hungary, and above all Russia and Kazakhstan, although *forced expulsions* after World War II and *massive emigration* to Germany in the 1980s and 1990s have depopulated most of these communities. (Metapedia: *German language*, emphasis mine)

Within these white nationalist discourses, Germans are often portrayed as the ultimate victims, which is consistent with the proliferation of Holocaust denialism on each of these platforms, and thus the article insinuates that endangerment of this dialect, even by other European languages, is indicative of the replacement of Germanic people.

White supremacists frequently elevate German and Germanic languages, as seen in a nauseating Metapedia article on Nordicism, where the article authors tie racial and linguistic supremacy with Germanic peoples.

> […] The concept of a reborn "Germanic" or "Nordic" race is a leitmotif of those who fear the downfall of the European race. Like in the first part of the nineteenth century, discussions on the origin and identity of the *Germanic peoples nation* are once again dominated by historians, *linguists*, and folklorists. […] The idea of a psychic unity of humankind has been exposed as a leftist political agenda. Every year thousands board rickety boats, hide in the backs of trucks, planes, and container ships, cross miles of barren desert on foot … All to get themselves to a land where they can be ruled by a racial group far distant from their own. While it's not clear what mix of genes/*language*/culture has permeated the Germanic space, we have to draw lines somewhere. We consider 'Teutonic' all the descendents of NW European *Germanic language speakers*, including all of the British Isles and their offspring-states, Scandinavia (with Finland), Germany, Austria, and Benelux. (We do not include South Africa, as even though it was built and maintained by Teutons, the reins of power have now been passed to others.) Though we are always told that the global 'South' wants to migrate to the 'North,' the reality is more nuanced. Hundreds of millions of people from the third world would like very much to live under the rule of a Teutonic people. (Metapedia: *Nordicism #History*, emphasis mine)

The discussion of using linguistics to reconstruct Proto-Germanic for some kind of Nordicist cause shows how the field of linguistics has been weaponized

by white nationalists. The appeal to some kind of modern 'Teutonic' people is a clear emblem of a delusional mythical past used to both define and justify the hegemonic position of various countries with Germanic official languages or lingua francas. Their exclusion of South Africa from this label is an apparent allusion to the end of apartheid and rise of a multiracial democracy in the country, which is viewed in the article as a loss for Germanic people.

We see the same kind of racism on GHC, where languages (as well as countries) are viewed as inherently white, which is the perceived source for their high value and demand across the world.

> All races want to breed with Whites, live in White countries, learn White languages, adopt our looks, our mannerisms, our NAMES. Love us, fuck us, kill us, live with us. Love/hate is a powerful combo, on the other hand all Whites want is some space to ourselves! (GHC: 17089)

This post also observes and justifies the hegemonic position of white people and then used against people of color.

3.3 Minority and endangered languages and whiteness

We see many instances where white nationalists view that nationalism, heritage, race, and languages are co-extensive, which is exemplified in the following post on GHC.

> Also useful for understanding the requirements of nationhood: common religion, common blood, common language, common culture (GHC: 290)

But critically they also see minority European languages as under threat, especially to larger languages or globalization.

> I would think that European nationalist organizations would be easier to create than American nationalist organizations. White Irish nationalist for Ireland White German nationalist for Germany White Dutch people for the Netherlands Thats not hard borders are very important even between European countries *To destroy or harm any culture is a crime against humanity and melting down all the smaller countries into one big country destroying there languages* and so on would just be wrong, we should work together yes we are all white and all brothers and our nations are sacred. (HSD: 12860520, emphasis mine)

Thus, white nationalists are concerned about minority languages to some extent, but when that's the case, it is for European minority languages. In doing so, these posters espouse white supremacists ideologies, but at the same time, portray white people as the victims.

We also see some concern among white nationalists on language and dialect endangerment, but again only European languages and dialects. We see this precise concern in the Metapedia article on *Constructed language*, though in the same breath, it still justifies the hegemony of national languages or lingua francas.

> Thus, for example, many dialects and small languages are gradually disappearing, but not due to competition from constructed languages, but due to competition from larger natural languages […] Even if arguing for that every language and dialect should be preserved, then in international communications the parties will still usually communicate using some particular language that both parties understand." (Metapedia: *Constructed language*)

White nationalists also lament the death of languages, but again only of European languages, and in this case, an ancient one, as conveyed with the initial adverbial.

> Unfortunately, all of the Eastern Germanic languages went extinct starting in the 4th century, and the last living language of this branch died in the late 18th century. (Metapedia: *Germanic languages*)

Concerns about language endangerment within white supremacist spaces likely parallel racist narratives of "white genocide," as exemplified in the following post.

> Yah it is sad to see and if the Irish dont quash it soon the last of the Kelts will be gone from the world." (HSD: 13852977)

Among white nationalists, there are pushes toward restoring and reclaiming this mythical past. As we see time and time again, their call for this historical and cultural reclamation entails the removal of anyone outside those who white supremacists define as their heritage.

> The Europeans Saxons, vikings, celts, normans, goths, vandals, romans, franks and many other European tribes need to come back to life, and reclaim there country, and culture before we no longer have a history or country we can call home. (GHC: 16802)

This sort of hearkening to a mythical past is constantly reinforced by white supremacist ideologies, as we see in the following post about Gaulish warriors:

> Ever noticed how they always say that about the Romans or Greek armies ... No one says "oh I wonder how many non Whites were on the Goth or Gaul armies" (HSD: 14060956)

This brings me to the principal case study of this paper. In one sub-forum on Stormfront Ireland, a poster discusses printing leaflets.

> I would like to get some of the leaflets printed and distribute them but sadly I have not got the funds. (HSD: 13642982)

The next post sampled in that subforum provides context for the event that the leaflets are being printed.

> With the help of God we will be congratulating an Irish WN Party on their election success in the not so distant future. 88 (HSD: 13842729)

When white nationalists, abbreviated in the post as "WN," discuss printing "leaflets" on Stormfront, they mean white supremacist literature. It is not clear which Irish election they are referring to, but posts in this corpus range from 2002 to 2017.

There are a number of subsequent posts within this subforum on the Irish election and politics. Eventually, another user brings up how to type in the Irish orthography, responding, at first, in an anglicized Scottish Gaelic.

> *Latha math dhuibh agus dè tha dol agad fhèin, a Phàdruig*[13] [...] *Aye, warriour of Éire*[14] ... whether you would like *fadas*[15] for *Irish Gæilge*, e.g., é, í, ó and the like, or *Scottish Gælick*, e.g., ì, ò, è, é, í, à, ù and the like ... try such, lad: "I. Start Menu to All Programs to Accessouries to System Tools to *Litrichean*[16] (I *canæ* re-call of what Litrichean tis called in English for my computer is in *Gàidhlig*,[17] but tis Script Lettres, Characters or the like, I reckon.) You shall know of where to go from there, *gille*.[18] or ... II. Change the language of

[13]Scottish Gaelic: *Good day to you (hon.) and what's going on with you, Patrick (voc.)*
[14]Irish: *Ireland*
[15]Scottish Gaelic and Irish: *diacritics*
[16]Scottish Gaelic: *letters*
[17]Scottish Gaelic: *Gaelic*
[18]Scottish Gaelic: *lad*

> your keyboard setting in your Control Panel under the Regional Optional and Languages theme. or … III. Make a visit here, lad … [19] or … IV. *Pray vnto Odin.* Many-a-luck to you, *gille.* (HSD: 13868014, emphasis mine)

The post on Stormfront, which code-switches frequently between English, Scottish Gaelic, and Irish, explains how to install Irish orthography on a keyboard, in response to someone printing leaflets for an upcoming election. This is one instance of white supremacists appropriating the language revitalization of Irish and Gaelic for the production of white supremacist literature.[20]

The "Pray vnto Odin" salutation is a nod to white supremacists' appropriation of Norse paganism (Pollard 2016) and Nordicism. In that vein, Wilson (2021: 249) noted that:

> Post-war fascists have also sought to bring about a revolutionary consciousness through the development of religio-political forms. Notable among them for its influence within the milieu, the quasi-Nordic Wotanvolk. […] Its racialism characterizes Wotansvolk as a "blood religion" in which "the folk" – and its purity – is central to the collective credo.

[19] Here, an external URL is provided.

[20] A reviewer for a conference presentation of this paper observed "the construction of Irish orthography as 'white' is surprising given that Irish immigrants in the US were previously not racialized as white." However, it is important to reiterate that this forum is a global venue for white supremacists. Many posts on Irish are specifically from the 'Stormfront Ireland' subforum. US history of Irish racialization is not appropriate lens for racism in Ireland and is arguably not even a good lens to understand contemporary racism in the US.

In fact, some white supremacists argue against distinguishing the relative whiteness of different nationalities, as is characteristic of historical anti-Irish, anti-Italian, and anti-Polish discrimination, but instead white supremacists in HSD appear to have mostly recast the debate between white and non-white, as seen in the following posts.

> I find it funny that people here spend their time trying to figure out if Spaniards, Ukrainians, Portuguese, Italians, and Russians are white enough. (HSD: 14666376)

> Meds are not superior, neither are Nordics, Germans are not better than Slavs, neither are Slavs better than Celts. Stop this BS. Our enemies want us ALL dead, regardless of which White tribe/region we belong to. (GHC: 1736)

To the extent that any white nationalist hold any white group as superior to another is usually a promotion of their own heritage rather than in disparaging another.

> Germanic, Scandinavian, Scottish! Best of the white race! Scottish temper with Germanic and Scandinavian build! Gotta love it! LOL 😂 (GHC: 16959)

> The Slavic race is superior!! (GHC: 831)

Wilson (2021) similarly discusses the appropriation and contortion of Wicca and Celtic Druidism by white supremacists.[21] The use of ⟨v⟩ in the word *vnto* is often a marker of white supremacy and Aryanism, as seen in the name of the now-defunct white nationalist group *Identity Evropa*.

The notion of someone responding in Scottish Gaelic to a post about an Irish election is an appeal to a broader pan-Celtic white supremacist solidarity. Other similar uses of European languages to appeal to a pan-European white supremacism is found on HSD.

> Why not have it so every flyer has all the White languages on it (in small print), yet have their native language in large print once on the top of the other translations, then once below it. Just to show all Whites should have one common goal. Unity. (HSD: 14417871)

This poster reinforces a notion of Europeanness which excludes people of color. The mention of both the local language and a pluralism of other European languages, the so-called "White languages," show how white nationalists racialize language.

The revitalization of the Scottish Gaelic language has been analyzed as a *culturally* nationalist project (McEwan-Fujita 1997). Similarly, Irish language revitalization has been associated with Irish nationalism and republicanism (Zenker 2012). However, the appropriation and weaponization of the revitalization of these two languages by white supremacists hasn't been well-documented and largely ignored.

[21]Many other Celtic elements are appropriated by white supremacists. For example, Celtic symbols, like the Celtic cross, have been famously appropriated by white supremacists, notably the KKK. Metapedia even acknowledges that the Celtic cross is "used by some nationalist groups stating a relationship with Celts" and "some more general White nationalist groups" (Metapedia, *Celtic cross*).

Moreover, we see Celtic mythology appropriated by white supremacists.

> I honestly don't care about this desert rat BS. I am a CELT! Heil LUGH! (GHC: 19938)

The poster exclaims pride in being a "Celt," and then uses the Nazi interjection with the Irish mythological figure Lugh. As such, the poster indexes their white supremacism through the proclamation of their pride of Celtic heritage.

Lastly, white nationalists list Celtic music among their preferences.

> […] Listen to Bluegrass, Celtic, Texas Swing, Rock a Billy, old time County, Southern Rock, and White Nationalist music! […] (HSD: 30383102)

> This filth is a sectarian British song not an Irish folk song the sort of scum who wrote and admire this filth have no place on a WN forum yet alone an Irish forum (HSD: 13857116)

Historically, some famous Irish language activists, notably Gearóid Ó Cuinneagáin, were fascists aligned with the Axis to resist the UK. Ó Cuinneagáin founded *Ailtirí na hAiséirghe* and used "language classes" as a guise to organize fascist meetings.

Another Celtic language, Breton (in Brittany), has received more attention in this regard. It is well-documented that Nazi Germany collaborated with fascist Breton groups in Brittany, like *Emsav*, during WWII. The Nazis promised support for Breton independence from France in exchange for an alliance with the Axis (Leach 2010). The alliance was not just a military and political one but was also philosophical as it adopted and adapted Nazi white supremacist ideology for a greater pan-Celtic solidarity and supremacy (Leach 2010).

Some early key Breton language revitalization activists were Nazi collaborators (Leach 2010, Boomgaard 2008). Some of these Breton language activists include Célestin Lainé, Olier Mordrel, and Roparz Hemon. Célestin Lainé led *Bezen Perrot*, the SS affiliated militia in Brittany, which was named after a Breton language activist and Nazi sympathizer Jean-Marie Perrot (Leach 2008). Lainé later fled to Ireland to avoid prosecution (Leach 2010). Olier Mordrel was a professor of Celtic languages and founder of the *Breton National Party*, which was later disbanded because of its connection to the Third Reich. Both Lainé and Mordrel broadened the racist notion of Nordicism to include all Celto-Germanic groups (Leach 2010). Perhaps the most influential to Breton activism and revitalization was Roparz Hemon, who was also accused of being a Nazi collaborator. He was sentenced to prison but instead exiled to Ireland. He continued to publish on Breton linguistics in Ireland and his grammar and dictionary are still considered authoritative within Breton linguistics.

In the aftermath of WWII, the Breton movement democratized over time and now largely embodies left-leaning ideologies, just like Scottish and Welsh nationalism,[22] which is fairly common among sub-state nationalism (Alessio 2015). Nevertheless, the history of the appropriation of Breton language activism by Nazi collaborators still casts a shadow over genuine progress of its current revitalization (Hornsby & Quentel 2013, Hornsby et al. 2021).

[22] In a more contemporary Welsh context, the now-defunct Welsh Defense League (WDL), sometimes called the Cymru Defense League, has received attention in the scholarship (Alessio 2015). Founded in 2009, it was an Islamophobic group whose leadership had ties to many different Neo-Nazi and Aryan white supremacist groups and it used violent means. Nevertheless, it did not grow much, as Welsh nationalism is largely left leaning, dominated by the Labour and Plaid political parties, and inherently tied to the Welsh language and opposed to creeping monolingual English-speaking groups. This contrasts with the WDL, who viewed Muslims and immigrants as threats. Indeed, the WDL and many of its successors have been more English-speaking and largely pro-British and Unionist (Alessio 2015).

3.4 Consequences

One poster makes explicit why language is important to white supremacists in their discussion of raising children, which reveals the sheer sinister nature of their goals.

> *Our youth are our future indeed! Just as Hitler's youth ours must be addressed early to the doctrines* in order to secure a foundamental inherent belief system which corressponds within a timely manner enabling them and thier thought processes to formulate and adjust to rudimentary belief functions incorporating all thier senses.They must become keenly aware of social agendas and political ramifications of action vs. inaction as it pertains to strategical logistics of a growing super power. *Home schooling if feesible* since, they must not only master the *social sciences* etc ... but als be able to be profiecent in the *arts* including *music theory*.Social skills and graces along w/*public speaking.Weapon, hunting, survival, building, self-defense, exsplosives/bomb making & deactivation, codes, languages, counterintelligence, interogation, re-programing/minds* and above all else always *behave as a little kid*! Gotchya-this can all be accomplished through father and *son/daughter communication* starting at a *young age* and progressing untill the late teen years *following dads clues/guidance*.Learned behavior-especially since children tend to *emulate the role model and mold them 'selves accordingly to thier parent* of whom they love very much and want to please. [...] (HSD: 14417601, emphasis mine)

This poster first begins with a sort of roadmap for white supremacists' parents raising their children, appealing to philosophies of hate propagated by Adolf Hitler and the so-called "14 words."[23] Recognizing the impressionability of children, this post discusses white supremacists' agenda of indoctrinating children with racist and violent ideologies early in their life. The poster suggests that much of this can be done in the home and therefore suggests home schooling, likely to exert total control over the instruction and prevent any influence from public school systems. The poster then lists a number of skills and expertise valued by white supremacists for their violent ends of white supremacy. However, many of these cannot be attained in white supremacist home-schooling environments since they involve socialization with others as well as skills that they can't learn from their parents. One of these are languages, disturbingly placed in an array of other skills like "weapons," "explosives," and "re-programming of minds."

[23] https://www.adl.org/resources/hate-symbol/14-words

The reinforcement of "behaving as a little kid" and everything thereafter highlights both the indoctrination and suppression of children while being taught these skills, all for the advancement of white supremacist causes. This seemingly involves infantilization of the children as they grow and into teenage and young adult years. The posts make clear that it is the goal of raising and indoctrinating new generations of white supremacists that will continue their legacies of violence and hatred.

Are they successful? There is considerable evidence that many are, when we consider the rising hate crimes and terrorism. Moreover, posts on Stormfront also confirm this.

> i am a 16yr old irish male dont worry lads i ll make sure ireland stays white and i hope i can count on help from my stormfront brothers around the world???? […] (HSD: 13864391)

This poster, in his commitment to "make sure Ireland stays white" is responding to calls for younger generations to continue on the legacy of white supremacism, violence, and hatred.

Why is there all this interest in languages? For one, learning languages are seen as opportunities in the spread of far-right ideologies, especially along the lines of heritage.

> By talking German the white Americans can better communicate with their white ethnic German cousins in Germany and bridge the gap. Even communication between English people and Germans may be improved – share political ideas in German. (GHC: 16042)

Further, a common use of international white supremacist spaces, like Stormfront, is for offers and requests for language and translation assistance from speakers of different languages, either for literature or in preparation of planned events and demonstrations. We can see that in the following post.

> You can ask to your french fellow in stormfront french part for a little help both of them speak english and french. (HSD: 30782282)

Different language communities have separate subforums on Stormfront, where these requests are made. Surprisingly, perhaps, there is also a subforum for Latin America on Stormfront, where one white supremacist referred another for help on translating literature.

> I've posted a thread in Stormfront Latin to help spread the WNPS and too see if we can get more translations. (HSD: 13463194)

This is indicative of a larger agenda of white supremacists. Southern Poverty Law Center (2022) has profiled individuals key to white supremacist movements, such as American white supremacist Gary Lauck, who translates and publishes white supremacist literature and propaganda and smuggles into countries which ban the material. He has translated material into nine languages and successfully smuggled it into twenty-nine other countries, upwards of 8 million pieces per year into Germany alone.

4 Conclusion

In this chapter, we have seen that white supremacists have sought to punish every departure of the constructed white straight male native speaker standard. Posters constantly reinforce a notion of Europeanness, based in a mythical history, but one who is under attack from and being replaced by non-white people who do not speak the national language. Some view minority and endangered European languages as also being the victims of this replacement.

A major research question in this paper was what languages do white supremacists use and view as "for whites only"? From this corpus, we see that white supremacists only accept Indo-European and Uralic languages. But for either, they only accept branches of the families native to Europe. Moreover, they only consider white "native" speakers, above all else, as legitimate speakers of European languages or white people who have heritage to the languages. They don't consider immigrants or people of color as legitimate speakers of these languages, and they don't accept immigrants or people of colors' own native or heritage languages. White supremacists view the speakers' race and ethnicity as deterministic of whether the speaker can and should speak a European language, as if it is only an inherited trait. However, they promote learning other languages, as long as they are languages native to Europe and as long as the learners are white. They also advocate for minority and/or endangered European languages and dialects, but in ways consistent with their white supremacist ideologies, such as connecting with some mythical heritage based in racism. Ultimately, their ideology has more to do with race of the speaker than even what language or variety is spoken.

For linguists, language instructors, and language activists, what can we do with this information? We need to monitor white supremacists' networks. We

need to understand how white supremacists appropriate and weaponize heritage languages, minority languages, endangered languages, and language revitalization, and that there are these dark subcultures among European language activists. Moreover, we need to be vigilant of how white supremacist consume and twist findings of comparative and historical linguistics for their racist means. We need to recognize the origins of these ideologies and how they seep into mainstream. We need to be cognizant of linguistic pseudoscience and folk theory, especially that which is rooted in racism, and actively expose it and combat it. We need to recognize that not everyone involved with language learning or linguistics share the same benign values or intentions. Finally, we need to include lessons of language or dialect discrimination into pedagogy, with a specific inclusion of minoritized varieties.

Acknowledgements

I'd like to thank Andrew Carnie, Sonja Lanehart, Joseph (Koodéik') Marks, Sebi Brinks, Benji Skoronski, five anonymous reviewers, and audiences at the University of Arizona Celtic Linguistics Group, the Linguistic Society of America Annual Meeting, Second Language Acquisition and Teaching, and the Hispanic & Lusophone Linguistics Working Group colloquium for their help in this project. Mistakes are my own.

References

Alessio, Dominic. 2015. The dragon is not always red: The extreme right and ultra-nationalism in Wales. *National Identities* 17(3). 289–309. DOI: 10.1080/14608944.2014.987658.

Anti-Defamation League. 2022. *Hate on display hate symbols database*. https://www.adl.org/resources/hate-symbols/search (5 June, 2025).

Arnstad, Henrik. 2015. Ikea fascism: Metapedia and the internationalization of Swedish generic fascism. *Fascism* 4. 194–208. DOI: 10.1163/22116257-00402002.

Beirich, Heidi. 2014. *White homicide worldwide: Stormfront, the leading white supremacist web forum has another distinction — murder capital of the internet*. Montgomery, AL: Southern Poverty Law Center.

Bergman, Harriet. 2021. Rising sea levels and the right wave: An analysis of the climate change communication that enables "the fascist creep". *Krisis: Journal for Contemporary Philosophy* 41(2). 2–18. DOI: 10.21827/krisis.41.2.37164.

Boomgaard, Michelle C. 2008. The rise of the miltant Bretonité. *National Identities* 10(3). 281–293. DOI: 10.1080/14608940802249890.

Bridle, Andrew. 2016. *The language of hate: A corpus linguistic analysis of white supremacist language*. New York: Routledge.

Byman, Daniel. 2022. *The global rise of white supremacist terrorism*. Oxford: Oxford University Press.

Calderón, Fernando H., Namrita Balani, Jherez Taylor, Melvyn Peignon, Yen-Hao Huang & Yi-Shin Chen. 2021. Linguistic patterns for code word resilient hate speech identification. *Sensors* 21(7859). 1–16. DOI: 10.3390/s21237859.

Carlson, Jedidiah, Brenna M. Henn, Dana R. Al-Hindi & Sohini Ramachandran. 2022. Counter the weaponization of genetics research by extremists. *Nature* 610. 444–447. DOI: 10.1038/d41586-022-03252-z.

Crenshaw, Kimberle. 1989. Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum* 1. 139–167.

Daniels, Jessie. 1997. *White lies: Race, class, gender and sexuality in white supremacist discourse*. New York: Routledge.

de Gilbert, Ona, Naiara Perez, Aitor García-Pablos & Montse Cuadros. 2018. *Hate speech dataset from a white supremacy forum*. 2nd Workshop on Abusive Language Online.

de Keulenaar, Emilie V., Marc Tuters, Ivan Kisjes & Kaspar von Beelen. 2019. On altpedias: Partisan epistemics in the encyclopaedia of alternative facts. *Artnodes* 24. 22–33. DOI: 10.7238/a.v0i24.3300.

de Keulenaar, Emilie V., Marc Tuters, Cassian Osborne-Carey, Danile Jurg & Ivan Kisjes. 2022. A free market in extreme speech: Scientific racism and bloodsports on YouTube. *Digital Scholarship in the Humanities* 37(4). 949–971. DOI: 10.1093/llc/fqab076.

Hartzell, Stephanie L. 2020. Whiteness feels good here: Interrogating white nationalist rhetoric on stormfront. *Communication and Critical/Cultural Studies* 17(2). 129–148. DOI: 10.1080/14791420.2020.1745858.

Hinton, Leanne, Leena Huss & Gerald Roche. 2018. Introduction: Language revitalization as a growing field of study and practice. In Leanne Hinton, Leena Huss & Gerald Roche (eds.), *The Routledge handbook of language revitalization*, xxi–xxx. New York: Routledge.

Hornsby, Michael, Nicole Dołowy-Rybińska, Joanna Chojnicka & Jeanne Toutous. 2021. The ideological foundations of Breton and lower sorbian language revitalization through education and their consequences for new speakers. *International Journal of Bilingual Education and Bilingualism* 8. 1–14. DOI: 10.1080/13670050.2021.1999900.

Hornsby, Michael & Gilles Quentel. 2013. Contested varieties and competing authenticities: Neologisms in revitalized Breton. *International Journal of the Sociology of Language* 2013(223). 71–86. DOI: 10.1515/ijsl-2013-0045.

Irvine, Judith T. & Susan Gal. 2000. Language ideology and linguistic differentiation. In Paul Kroskrity (ed.), *Regimes of language*, 35–84. Santa Fe, NM: School for American Research.

Kennedy, Brendan, Mohammad Atari, Aida Mostafazadeh Davani, Leigh Yeh, Ali Omrani, Yehsong Kim, Kris Coombs Jr., Shreya Havaldar, Gwenyth Portillo-Wightman, Elaine Gonzalez, Joseph Hoover, Aida Azatian, Alyzeh Hussain, Austin Lara, Gabriel Cardenas, Adam Omary, Christina Park, Xin Wang, Clarisa Wijaya, Yong Zhang, Beth Meyerowitz & Morteza Dehghani. 2022. Introducing the Gab hate corpus: Defining and applying hate-based rhetoric to social media posts at scale. *Language Resources and Evaluation* 56(1). 79–108. DOI: 10.1007/s10579-021-09569-x.

Kroskrity, Paul V. 2021. Covert linguistic racisms and the (re-) production of white supremacy. *Journal of Linguistic Anthropology* 31(2). 190–193. DOI: 10.1111/jola.12307.

Leach, Daniel. 2008. Bezen perrot: The Breton nationalist unit of the SS, 1943-5. *e-Keltoi: Journal of Interdisciplinary Celtic Studies* 4(1). 1–38.

Leach, Daniel. 2010. "A sense of Nordism": The impact of Germanic assistance upon the militant interwar Breton nationalist movement. *European Review of History–Revue européenne d'histoire* 17(4). 629–646. DOI: 10.1080/13507481003743559.

Lui, Marco & Timothy Baldwin. 2012. Langid. Py: An off-the-shelf language identification tool. In *Proceedings of the acl 2012 system demonstrations*, 25–30. Jeju Island, Korea.

McEwan-Fujita, Emily. 1997. Scottish Gaelic and social identity in contemporary Scotland. *Replika* 8. 195–156.

Oeberst, Aileen, Jort de Vreeze & Ulrike Cress. 2018. The norm of neutrality in collaborative knowledge construction. In Olga Zlatkin-Troitschanskaia, Gabriel Wittum & Andrea Dengel (eds.), *Positive learning in the age of information: A blessing or a curse?*, 209–220. Wiesbaden, Germany: Springer VS. DOI: 10.1007/978-3-658-19567-0_12.

Pollard, John. 2016. Skinhead culture: The ideologies, mythologies, religions, and conspiracy theories of racist skinheads. *Patterns of Prejudice* 50(4-5). 398–419. DOI: 10.1080/0031322X.2016.1243349.

Rosa, Jonathan & Nelson Flores. 2017. Unsettling race and language: Toward a raciolinguistic perspective. *Language in Society* 46. 621–647. DOI: 10.1017/S0047404517000562.

Southern Poverty Law Center. 2019. *It's not going well for Gab*. https://www.splcenter.org/news/2019/02/08/weekend-read-it%E2%80%99s-not-going-well-gab (5 June, 2025).

Southern Poverty Law Center. 2022. *Extremist Files*. https://www.splcenter.org/fighting-hate/extremist-files/groups (5 June, 2025).

Wilson, Andrew F. 2021. "Our community could start our own traditions": The commingling of religion, politics, and the folkloresque in a far-right groupuscule. In Matthew Cheeseman & Carina Hart (eds.), *Folklore and nation in Britain and Ireland*. New York: Routledge.

Zenker, Olaf. 2012. On prophets, godfathers, rebels, and prostitutes: Distributed agency in the Irish language revival of Northern Ireland. *Zeitschrift für Ethnologie* 137. 23–45.

Chapter 15

France's war on Breton diacritics: An incomprehensible obstinacy

Stefan Moal
Université Rennes 2

The government of the French Republic has refused to sign the European Charter for Regional and Minority Languages. In this chapter, I review a series of legal decisions and official memos that exemplifies this bias against regional and minority languages as expressed through restrictions on the use of orthographic diacritics not found in standard French. Using non-French diacritics is considered to be an acknowledgment of the right to use a language other than French when dealing with the government.

1 Introduction

Those looking at France from the outside might not see it as a linguistically homogeneous state, but rather as a union of different communities and peoples. However, the very notion of "minority" does not actually exist in France. While almost all other European states have ratified the Framework Convention for the Protection of Minorities[1] – now virtually a precondition for entry into the European Union – France has not. The repeated rejection of the European Charter for Regional and Minority Languages[2] (signed in 1999 but never ratified) even calls into question France's place in democratic Europe, since it remains one of the very few European Union member states to retain provisions privileging a single official language in its constitution. The only constitutional reference to

[1]https://www.coe.int/en/web/minorities/at-a-glance
[2]https://www.coe.int/en/web/european-charter-regional-or-minority-languages/text-of-the-charter

Stefan Moal. 2025. France's war on Breton diacritics: An incomprehensible obstinacy. In Andrew Carnie, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond & Luis Irizarry (eds.), *Foundational approaches to Celtic linguistics*, 527–553. Berlin: Language Science Press. DOI: 10.5281/zenodo.15654885

other languages in its territory was, until recently, Article 75-1 ("Regional languages belong to the heritage of France") in keeping with historical and archival thinking. Of course, actual use of these languages in everyday life should be encouraged instead.

On May 21, 2021, the Breton MP Paul Molac proposed the law "on the protection of the heritage of regional languages and their promotion"[3], but the Constitutional Council ruled out two of its major provisions as contrary to Article 2 of the Constitution ("The language of the Republic is French"): (1) immersive teaching, carried out during a large part of school time in a language other than French, like in Diwan schools for Breton; (2) the use of diacritics from regional languages on civil status documents (Conseil Constitutionnel 2021).

Using non-French diacritics is considered to be an acknowledgment of the right to use a language other than French when dealing with the government. This chapter examines the details of the latter decision. After reviewing the definition of diacritics, especially in the French context of the "Taubira" memo, I discuss the various court cases, from the initial Catalan lawsuit to the particularly infamous Fañch affair, currently ongoing in Brittany. In passing, I point out a few positive outcomes, notably for placenames, and we will see how parents who plead for the correct spelling of their children's names both in regional and in foreign languages have found a common cause to fight for. At the time of writing (summer, 2024), the situation remains deadlocked, and the current political turmoil in France is not conducive to a rapid resolution. Although the ban on diacritics is not as serious as the ludicrous French transcriptions of Breton placenames in the past (e.g., the infamous translations of *kroazhent* (crossroads) as *croissant*)[4], it can still be concluded that it is a symptom of a deeply rooted trend in France to demean so-called "regional languages".

2 Definitions

Let us start by reviewing the definition of "diacritic", borrowed from the Greek διακριτικός: "which makes it possible to distinguish". A diacritic is a mark added to a letter of the alphabet to indicate a modification of its pronunciation or distinguish the word that includes it from another, homonymous word. It can be

[3]https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000043524722

[4]These names endured much longer, and still do in some cases, than those given during the short-lived anti-religion/anti-monarchy renaming craze of 1793–1794: Montagne-sur-Odet (instead of Kemper-Korantin/Quimper-Corentin), Pont Libre (Pontekroaz/Pont-Croix), Port de l'Égalité (Porzh-Loeiz/Port-Louis) or Port Nazaire (Sant-Nazer/Saint-Nazaire), to cite just a few examples in Brittany.

placed above, below, in front, behind, inside or across the letter to which it is attached. In French law, the administrative doctrine relating to the use of diacritics in civil status records is based on several texts (De Bailliencourt 2018). The decree of Thermidor 2, Year II (July 20, 1794) requires that all public deeds be written in the French language throughout the territory of France. The decree of Prairial 24, Year XI (June 13, 1803) applies to all departments and all countries under Napoleon Bonaparte's Consulate and defines that "where the practice of drawing up the said [public] acts in the language of these countries has been maintained, [they] must all be written in the French language", although the translation into "the idiom of the country" may be written "halfway down the French minute". Article 2, paragraph 1 of the Constitution of October 4, 1958, resulting from constitutional law no. 92-554 of June 25, 1992, states that "the language of the Republic is French". Finally, Article 1 of Law no. 94-665 of August 4, 1994, on the use of the French language, states that "the language of the Republic under the Constitution, the French language is a fundamental element of the personality and heritage of France" and "is the language of education, work, trade and public services".

Based on these texts, the State requires that civil status records, which hold authentic value, be written in French and obey French spelling rules. The memo dated July 23, 2014 (Bulletin Officiel, Ministère de la Justice 2014) relating to civil status, known as the "Taubira memo" states that "only the Roman alphabet may be used" and that "the only diacritic marks allowed are dots, umlauts, accents and cedillas as they appear attached to vowels and consonants authorized by the French language". It further states that: "While Convention n°14 of the International Commission on Civil Status (ICCS) on the indication of surnames and first names in civil status registers recognizes foreign diacritic marks, it should be noted that this convention has not been ratified by France. The vowels and consonants accompanied by a diacritic used in French are: <à, â, ä, é, è, ê, ë, ï, ô, ö, ù, û, ü, ÿ, ç>. These diacritics can be used on both upper- and lower-case letters. The ligatures <æ> (or <Æ>) and <œ> (or <Œ>), equivalent to <ae> (or <AE>) and <oe> (or <OE>), are accepted in the French language. Any other diacritic attached to a letter or ligature cannot be used to establish a civil status record".

These rules ignore the principle that parents should be free to choose their children's first names: parents can do so and may use non-traditional spellings, provided they only include the diacritic letters and ligatures of the French language mentioned above. This provision excludes: <ñ> in Basque and Breton; <ò> in Creole and Occitan; <á, í, ó, ú>, in Occitan, Catalan and Corsican; *tārava* (literally: "bar") in Polynesian languages. The simplification of the rules and procedures for changing first and last names, by the law of November 18, 2016, "to

modernize justice in the 21st century" (Articles 60 to 61-4 of the Civil Code) have done nothing to change this state of affairs.

Yet the presence or absence of these diacritics are anything but trivial for a correct pronunciation in the above-mentioned languages. The <ñ> represents the palatal nasal consonant in Basque, while in Breton it has been used since 1907 to nasalize the preceding vowel – without the <n> itself being pronounced. In Occitan and Catalan, the acute accent indicates the place of the tonic stress or indicates that a diphthong should be cut. In Marquesan, Tahitian, Wallisian and Futunian, all Polynesian languages, the *tārava* lengthens the vowels <ā, ē, ī, ō, ū>. Kanak languages (in Kanaky/New-Caledonia), meanwhile, are spelled with the diacritic marks found in French.

In its decision no. 2021-818 DC of May 21, 2021, on the law on the protection of the heritage of regional languages and their promotion, the French Constitutional Council declared Article 9, which stipulates that the diacritic marks of regional languages are authorized in civil status records, to be unconstitutional (Conseil Constitutionnel 2021). This decision brought to a halt a rather chaotic evolution in the authorization or use of these diacritics to record names in French civil status registers. This process, which has been punctuated for a quarter of a century by a series of legal proceedings marked by numerous twists and turns, has mainly concerned users of certain regional languages – namely Basque, Breton, Catalan and Occitan – but now also concerns French citizens of Spanish- or Portuguese-speaking descent, who are also prohibited from registering the tilde in their surname or first name. It is this process that will be discussed here, as well as the possible outcomes of the decision made by the so-called "Sages" ("Wise Men/Persons").

3 A memo (mis)named Taubira

When Christiane Taubira was still a member of parliament, representing the South American French department of Guiana, situated between Suriname and Brazil, she published an article on her blog hosted on the *Médiapart* information website, entitled "Langues et Cultures 18 mars 2012" (Taubira 2012) to commemorate the fifth anniversary of the entry into force of the UNESCO Convention on the Protection and Promotion of the Diversity of Cultural Expressions. It may therefore come as a surprise that Christiane Taubira never commented on the July 23, 2014 memo bearing her name.

She is presumably much more honored to see her name associated with the May 21, 2001 law "recognizing the slave trade and slavery as a crime against

humanity", or the May 17, 2013 law "opening marriage to same-sex couples", than with a text that bears her name only because she was then *Garde des Sceaux* (Minister of Justice). Herself a fluent and proud speaker of Guianese (French-based) Creole, she then reminded her readers that "minority languages are often majority languages in their own right. [...] They can become official when the people decide to do so and history leads them to do so. This is the case of Haitian Creole in Haiti and the Seychelles. She asserted that "languages don't just convey words, they carry imaginations, universes", recalling that "the original diversity of French cultures was sacrificed on the altar of a single, indivisible Republic, backed by a Nation that believed it could only prosper through uniformity. [...] The Nation waged war on regional languages and cultures".

The Guianese representative went on to describe as "brutal" the term *baragoin* (gobbledygook, gibberish) used by a President of the Roazhon/Rennes Court of Appeal who had adjourned a session on the grounds that two defendants were speaking Breton. As far as regional languages are concerned, she considered that "the Toubon Law of 1994, blinded by its battle with English and anglicisms, had almost destroyed everything". Concluding that "the issue is not linguistic, it's political". Insisting on the importance of transmission, she invited the readers of her blog to ask "Paco Ibanez[5] what he received from his grandmother in Basque".

Ibañez happens to also be the name of the founder, in 2017, of a Facebook group called *Touche pas à mon N tildé* (Don't touch my N tilde), rebranded in 2024 *Mon N tildé, ma bataille* (My N tildé, my battle) which as of July 2024 had 719 members and federated all initiatives aimed at the recognition by French civil status of all diacritic marks currently banned. The long legal battle over diacritics goes back much further, however, since it began at the very start of the 21st century.

4 The *Martí* case in North Catalonia, starting point of a long judicial process

In the early 2000s, the first case concerning the Catalan first name *Martí* gave rise to a lengthy and highly-substantiated dispute, which at the same time provided an opportunity to set out the facts of the legal problem (Esteban 2016). When his

[5]Francisco Ibáñez Gorostidi (born 1934 to a Valencian father and a Basque mother) is an anarchist singer, fluent in Basque and Catalan in addition to French and Castilian. Despite Christiane Taubira's words in strong favor of "regional languages", which we have no reason to doubt are sincere, one may wonder whether the double absence of an acute accent on the <a> and tilde on the <n> in the singer's surname wasn't somehow foreshadowing the 2014 memo bearing her name, and the judicial and constitutional twists and turns that followed.

son was born in 1998, Alà Baylac-Ferrer never imagined that ten years later he would file, and lose, an appeal with the European Court in Strossburi/Strasbourg to register his son's name, *Martí*, correctly spelled. Ever since, the third child of the Baylac-Suarez family has been refused registration of the accent on the <í> in various courts, as this character does not belong to the French alphabet: in France, Martí is called Marti, therefore the stressed syllable is not the same and as a result the pronunciation is incorrect.

Baylac-Ferrer is a Catalan language lecturer at the University of Perpinyà/Perpignan and one of the main promoters of the language in North Catalonia (department of Pyrénées-Orientales, South of France). The Baylac-Suarez couple have another son with a first name also of Catalan origin, Joan, which posed no problem as all its characters appear in the alphabet considered legally French. When Alà Baylac-Ferrer signed the papers at the registry office a few days after Martí's birth, he noticed that the accent on the newborn's name had disappeared. This policy of just removing the offending diacritic was plausibly meant to keep the parents from noticing and causing issues. Who knows how many busy and tired new mothers might have missed this and ended up with the incorrect name for their child as a result? The registrar's justification was that the character didn't appear on the computer keyboard, to which Alà Baylac-Ferrer replied that all he had to do was add the accent in pen to the paper. However, the civil servant refused, fearing that doing so would constitute a forgery.

It all appears to be a matter of conflicting norms: on the one hand, Article 57 of the French Civil Code guarantees parents' freedom to choose their child's first name, stating that "the child's first names are chosen by the father and mother"; on the other hand, the 1999 general instruction on civil status, based on the principle of mandatory use of French enshrined in Article 2 of the Constitution and the law of August 4, 1994, states that "documents must be drawn up in French [...]. It follows in particular that the alphabet used must be [...] the only alphabet in use for writing the French language". This alphabet includes certain diacritic marks, but excludes others that are used to write some of the "regional languages" spoken in France, including Catalan.

It was first the high court (*Tribunal de Grande Instance,* Perpinyà/Perpignan, February 13, 2001) and then the court of appeal (*Cour d'Appel*, Montpelhièr/Montpellier, November 26, 2001) that had the task of resolving the contradiction between the aforementioned provisions. In both cases, the judges ruled that the freedom to choose a first name established by the French Civil Code could not override the obligation to use French and its alphabet in civil status documents. The registrar must agree to transcribe first names in "regional" or "foreign" languages, but these must be spelled *à la française* which excludes diacritic marks

not used in French. The first name *Martí* should therefore be spelled *Marti*.

The next step was the last possible one within the French legal framework, namely the Paris Court of Cassation, but its decision was yet another setback (March 2, 2004). Martí's parents, believing that their fundamental rights had been violated and noting that their domestic remedies had been exhausted, took their case to the European Court of Human Rights in Strossburi/Strasbourg. They invoked the principle of freedom to choose one's children's first names, guaranteed by the French Civil Code, as well as the right to respect for one's private and family life, enshrined in Article 8 of the European Convention for the Protection of Human Rights (Ravasi 2017).

However, the European Court in Strasbourg rejected the Baylac-Suarez family's appeal in application of the principle of proportionality, stating that freedom of linguistic choice of a first name "does not form part of the matters governed by the Convention" and that member states thus enjoy a wide margin of appreciation as regards rules on surnames and first names (European Court of Human Rights 2012). The harm would have to be particularly significant and undermine the child's personal identification to justify a violation of Article 8 of the Convention in relation to the linguistic regulation of first names.

Again, the ECHR validated the principle of acceptance of non-French first names by civil registries, provided they are spelled according to the French alphabet. It considered that the refusal to spell a first name in Catalan does not constitute an unjustified and disproportionate infringement of the parents' right to a private and family life (Article 8 of the European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR), nor discrimination on the grounds of belonging to a national minority (Article 14 ECHR), nor an infringement of the parents' right to a fair trial (Article 6 ECHR). It also added that "the justification put forward by the Government, namely linguistic unity in relations with the administration and public services, was imperative [...] and proved to be objective and reasonable."

5 Fañch, "the name that shakes the Republic" (Rouz 2020)

The second major episode in the long-running controversy[6] surrounding diacritics in names took place almost 20 years after the Martí affair began, far from the shores of the Mediterranean, in Brittany. On May 11, 2017, a little boy was born in Kemper/Quimper and his parents, Lydia Fuzier and Jean-Christophe Bernard, living in the nearby town of Rosporden, chose to name him *Fañch*, after one of his

[6]See Appendix 1 for a timeline of the Fañch affair rulings.

grandfathers, Fañch Cotten from Elliant. This grandfather was in fact registered as *François*, historically one of the most popular male first names in Lower-Brittany, and which, ironically, originally meant 'Frank' and by extension 'French'. Men called *François* in Lower-Brittany are most often referred to by various Breton versions of the name: *Frañsez, Frañsa, Saig, Soaig, Feñch*, and *Fañch*. Occasionally, there would even be a *Saig* and a *Fañch* among the siblings of one same family, both officially called *François*. The newborn's parents told the Kemper/Quimper registry office that *Fañch* should be spelled with a tilde over the <ñ>, but the officer initially refused to do so, transposing *Fañch* into *Fanch* on the grounds that the tilde is not a diacritic used in the French language. The parents contacted the town hall of Kemper/Quimper, and the first deputy mayor decided to have the tilde added to the birth certificate. She invoked Article 75-1 of the French Constitution, Article 57 paragraph 2 of the French Civil Code and a 1996 ruling by the European Court of Human Rights, while appealing to the administrative authorities for understanding. Anticipating possible complications, Fañch's parents immediately had an identity card and passport issued in their son's name with the tilde over the <ñ>, which didn't pose any technical problems for the services concerned. The Roazhon/Rennes Public Prosecutor, who is responsible for ensuring that civil registrars comply with the law, challenged the decision in court and opened the legal proceedings. At the beginning of July 2017, the parents were summoned to appear before the Kemper/Quimper High Court with a view to a "request for rectification of a civil status record, or of declaratory or suppletive civil status judgments" (Feltin-Palas 2017).

The main arguments of the public prosecutor's office were as follows: The law of Thermidor 2, Year II, stipulates that public documents must be written in the French language. In accordance with Article 2 of the French Constitution, "the language of the Republic is French", the Constitutional Council (Conseil Constitutionnel 2021) ruled in 2001 that the use of French was mandatory for legal entities governed by public law and private law in the exercise of a public service mission. In 2011, the same Constitutional Council ruled that Article 75-1 of the Constitution, which since 2008 states that regional languages are part of France's heritage, did not confer any rights or freedoms that could be enforced by individuals. In the Martí case, the ECHR ruled in favor of France in 2008, and even deemed "objective and reasonable" the justification put forward by the French government for the need for linguistic unity in relations with the administration and public services (European Court of Human Rights 2012). A 2014 memo lists exhaustively the diacritic marks admitted by civil status, exclusively those authorized by the French language, specifying that any other sign cannot be retained.

Convention 14 of the International Commission on Civil Status (ICCS)[7] on the indication of surnames and first names in civil status registers recognizes foreign diacritic marks, but France has not ratified that convention.

Fañch's parents also put forward their own arguments. They were astonished that, in 2017, France continued to refer to the law of Thermidor 2, Year II, passed at the height of the *Terreur* just a few days before the fall of Robespierre, which also stipulated that any civil servant who failed to write in French would be brought before the criminal court of his residence, sentenced to six months' imprisonment and dismissed. They pointed out that the Toubon Law of 1994 on the use of the French language was intended to defend French against the hegemony of English, not against regional languages, since Article 21 stipulates that it applies without prejudice to legislation relating to regional languages and does not oppose their use. They were not asking for their son's birth certificate to be fully written in a regional language but simply for his first name to be spelled in Breton. The tilde has also existed in French, as demonstrated by Bernez Rouz, President of the Cultural Council of Brittany. It is even found in the 1539 Villers-Cotterêts ordinance by King François 1st, which in its Article 111 proscribed the use of Latin in all judicial documents and imposed instead the use of the "French mother tongue" – arguably not just French, but any of the languages spoken natively in the kingdom at the time. They challenged the Constitutional Council's interpretation of Article 75-1, according to which the introduction of regional languages into the fundamental law does not give rise to any rights or freedoms enforceable by individuals: if these provisions have no normative scope, then they are purely declarative, even decorative. While the 1999 general instruction on civil status and the 2014 memo stipulate that only the Roman alphabet may be used, the <ñ> does belong to this so-called "Roman" alphabet (in fact: *Latin* alphabet) as demonstrated by its existence in several Romance languages. Accepting <ñ> poses no technical problem since the sign appears on all computer keyboards. According to Article 57 of the French Civil Code, "the child's first names are chosen by his or her father and mother" and no mention is made of a ban on spelling first names in regional languages. Article 8 of the European Convention on Human Rights states that everyone has the right to respect for his private and family life. The emotional choice of a first name falls within the scope of the parents' private life. The same Article adds that there shall be no interference by a public authority with the exercise of this right except in the interests of national security or the protection of health, which are certainly not

[7] An intergovernmental organization whose aim is to promote international cooperation in civil status matters and to improve the functioning of national civil status services.

threatened by a tilde in a first name. Article 14 of the same Convention stipulates that the enjoyment of the rights and freedoms set forth shall be secured "without discrimination on any ground such as sex, race, color, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status." By refusing to accept the traditional spelling of one of the languages in its territory, the French state would be in contradiction with the European Convention on Human Rights.

In its ruling of September 13, 2017, the court annulled the correction made by the registrar, thus reinstating the spelling *Fanch*, without the tilde. The parents had not hired a lawyer to prepare their defense, as they probably didn't imagine such excesses on the part of the public authorities, especially as the Kemper/Quimper civil registry had already, very quickly, vindicated them (Blanchet 2017). Unhappy with this court injunction, but very determined, they took their case to the Roazhon/Rennes Court of Appeal. Local authorities in Brittany supported them: after the town of Kemper/Quimper, the Regional Council of Brittany unanimously adopted a vow on October 13, 2017 – the *Front National* party not having taken part in the vote – just like the Penn-ar-Bed/Finistère Departmental Council the following week, also unanimously. Nathalie Appéré, mayor of Roazhon/Rennes, sent a letter to the Minister of Justice, Nicole Belloubet, invoking the French Constitution according to which "regional languages are part of France's heritage", as well as the freedom to choose a first name as long as it is not contrary to the child's interests (Civil Code, Article 57).

Fourteen MPs from Brittany belonging to President Macron's majority, including the President of the National Assembly, sent a letter to the Minister of Justice, dated September 22, 2017, in which they asked her to kindly amend this memo so that spellings of first names in regional languages can be authorized, even if they use particular diacritics. They expressed their astonishment that a memo should prevail over the French Civil Code and Constitution, and that two very ancient texts were dug up: law no. 118 of Thermidor 2, Year II (July 20, 1794) and the decree of Prairial 24, Year XI (June 13, 1803). As of February 2024 though, nothing has changed in the ministerial memo. Seen from abroad, particularly from Spain but not exclusively, there was a great deal of incomprehension expressed: "this information is so surreal that you'd think it was fake news" commented *El Periódico de Catalunya*, for example, using dark humor in its headline alluding to the guillotine: "France beheads the ñ". The Barcelona daily pointed out that, despite appearances, this was "not an isolated case, nor a small unimportant detail", since "the stubbornness to impose French as the sole language is such that France entered a reservation when ratifying the Convention on the Rights of the Child [in 1990]", stating that Article 30 of the Convention, on the cultural,

religious and linguistic rights of minorities, did not apply to the territory of the Republic "where there are no minorities, whether cultural or linguistic" (Courrier International 2018).

6 Fañch wins on appeal then in cassation

On October 19, 2018, the court in Roazhon/Rennes ruled against the Kemper/Quimper court and decided to grant the petitioners' request, accepting the legality of registering the first name *Fañch* in its Breton spelling. Fañch's parents' lawyer, Mr Kerloc'h from the Naoned/Nantes bar, argued for their legal right to freely choose their child's first name, and warned against possible discrimination. The Advocate General, for his part, relied on the letter of the July 23, 2014 memo on civil status and pointed out that <~> is not included among the diacritic marks of French. Between a claimed right to difference and the strict observance of the law as currently established, the Roazhon/Rennes Court of Appeal's ruling therefore arbitrated this time for the former. The argument followed by the Court of Appeal reflects a conciliatory spirit and sets the debate on a linguistic level, since the Court proceeds to an *in concreto* control of the use in the French language of the tilde of the letter <ñ>. It notes that the <ñ> is used in French, as it "appears several times in the *Académie française* dictionary, in the Larousse dictionary and in the Petit Robert", and it is because it considers the <ñ> to be a French-language diacritic mark that the Court of Appeal authorizes the transcription of the first name *Fañch* in its Breton spelling (Gicquel 2019).

The court also referred to precedents where the use of <ñ> in official documents had been allowed by the administration, including cases in Roazhon/Rennes (2002) and Paris (2009) where the same name *Fañch* had been transcribed by officers of civil status in its proper Breton spelling. It also cited presidential decrees appointing senior civil servants whose surname is spelled using the letter <ñ>: notably Laurent Nuñez, appointed Secretary of State to the Minister of the Interior precisely in the midst of the Fañch affair. The judge was therefore using linguistics to call into question the provisions of the memo of July 23, 2014, by reinstating the letter <ñ> in the corpus of signs used in French. At this time, the matter seemed settled, after eighteen months of proceedings for Fañch's parents.

The president of the Cultural Council of Brittany, Bernez Rouz, published a book about the Fañch affair (Rouz 2020), in which he recalled that the tilde was used to mark nasals for centuries in French, even in ancient texts emblematic of the official status of French. The above-mentioned Villers-Cotterêts ordinance contains several tildes, for example in the word "*cõtenus*" (*contenus*). The State

recently relied on this pre-revolutionary ordinance to oppose any official use of "regional languages". In another official document from 1567, King Charles IX also used the word *cõsiderations* (considerations). Bernez Rouz's efforts to demonstrate that <ñ> is indeed part of the French orthographic tradition are of course laudable, but the spelling of a native language like Breton could also simply be defended and illustrated for itself.

As early as November 22, 2018, the Roazhon/Rennes public prosecutor's office lodged an appeal in cassation (abrogation), in a statement issued by Attorney General Jean-François Thony: "as the texts stand, the tilde is not recognized as a diacritic mark in the French language. It therefore seems necessary, in view of the possible national repercussions of the aforementioned ruling, to submit to the Court of Cassation the question of the use of the tilde in a first name". On October 17, 2019, the Court of Cassation rejected the appeal lodged by the Public Prosecutor of Roazhon/Rennes against the ruling of the Court of Appeal, and this time definitively closed the debate as far as little Fañch Bernard was concerned. His parents expressed their satisfaction that "no one would be able to take it away from him" henceforth. Picking up on a dispatch from the Agence France Presse, a number of major media outlets reported on this latest episode, including Le Monde and The Guardian.

This was however a victory by default, since it was obtained only on the grounds of a formal legal error (Le Monde 2019). The Roazhon/Rennes public prosecutor's office lodged its appeal solely against the parents, forgetting to do so "against in the capacity of legal representative of the child", which enabled the Court of Cassation, which does not judge on the merits, to declare the Rennes public prosecutor's appeal inadmissible. The July 23, 2014 memo on civil status records is certainly not abolished as a result of the Court of Cassation ruling. Although Mr Kerloc'h believed that it would be difficult to refuse the tilde to another child, the Brest public prosecutor actually announced to all the mayors in his court's jurisdiction that the problem remained unresolved, precisely because the Court of Cassation had not ruled on the merits and summoned them to notify him of any new birth of a *Fañch* with a tilde.

7 Four more Fañch affairs to date after the initial case

As the Brest public prosecutor had suspected, a second Fañch case appeared in Montroulez/Morlaix on November 18, 2019, just one month after the Court of Cassation rejected the appeal in the first Fañch case. It is also with the correct Breton spelling that the parents of the new little Fañch had his first name registered by the Montroulez/Morlaix civil registry, but the public prosecutor ordered

the removal of the tilde from this transcription in the birth certificate, again referring to the memo of July 23, 2014. This family also started a legal procedure but they did not wish to be exposed in the media and therefore did not communicate openly. They were represented and defended by the same lawyer, Mr Kerloc'h, and also supported by *Skoazell Vreizh*, an NGO founded in 1969, originally to help the families of "Breton political prisoners".

Little Awen Fañch was born on April 15, 2020, at Gwengamp/Guingamp hospital. His parents chose the middle name of his paternal grandfather, *Fañch* and filled in the civil registration form with a tilde over the <ñ>. A week after the maternity hospital forwarded the file to the municipality of Pabu, the family's home town, the parents noticed that the tilde did not appear on the birth certificate and first thought this was a mistake, like in the above-mentioned Baylac-Ferrer case. In fact, the town hall secretary had only done her job by complying with a memo from the Sant-Brieg/Saint-Brieuc public prosecutor. The family didn't hold this against her, but didn't intend to leave this "Fañch III" at that either, so they contacted *Skoazell Vreizh*. "While our initial approach was not at all militant, but rather a family affair", explains the mother, "from now on, if we have to be militant to ensure that the Breton language and our traditions are respected, we're ready to be militant for ourselves and for anyone else it might concern in the future" (Le Fur 2020). On May 4, the mayor of Pabu drew up a new birth certificate in which he had the tilde reinstated, contrary to the memo from the Sant-Brieg/Saint-Brieuc public prosecutor, who to date has not commented on the matter. If he were to object, however, Awen Fañch's family would go to court to try and win the case, because, adds the mother with a certain good sense: "the Breton language has precise rules, its own literature and a whole wealth of riches. I can't see why these rules should be ignored and considered a trifle, while one tries so hard to respect the rules of French".

In June 2023, the mayor of An Oriant/Lorient, Fabrice Loher, signed the birth certificate of yet another little Fañch, with a tilde over the <ñ>, even though the registry office had initially said no. However, after the summer, the public prosecutor of An Oriant/Lorient, Stéphane Kellenberg, sent a letter to the parents requesting that the tilde on the <ñ> be removed (James 2023). In this letter, he refers to the Constitutional Council's decision of May 21, 2021 "concluding that Article 9 of the law on the protection of regional languages and their promotion is unconstitutional". The letter states that "in its decision, the French Constitutional Council recalls, in accordance with Article 2 of the French Constitution, that the use of French is mandatory for legal entities under public law and for persons under private law in the exercise of a public service mission. Private individuals may not claim, in their relations with administrations and public services, a right

to use a language other than French, nor be compelled to do so." In this decision, the French Constitutional Council ruled that "by providing that entries in civil status records may be written using diacritic marks other than those used to write the French language, these provisions recognize the right of individuals to use a language other than French in their dealings with administrations and public services. They therefore fail to meet the aforementioned requirements of Article 2 of the Constitution."

The parents, Etienne Pichancourt and Mélissa Yana, thought that since the mayor of An Oriant/Lorient had validated the name, and Fañch had received his identity card, his tilde couldn't be taken away from him. The An Oriant/Lorient public prosecutor's office clearly didn't want to take this into account and decided to rely solely on the decision of the French Constitutional Council. "Insofar as this decision by the Constitutional Council, an obviously superior jurisdiction, is opposed to the case law of the Roazhon/Rennes Court of Appeal, I can only legally proceed with the administrative rectification of the purely material error vitiating the spelling of the first name attributed to your child, under the terms of Article 99-1 of the Civil Code, by giving you the necessary instructions for the deletion of the tilde appearing wrongly in this civil record, since this diacritic sign is non-existent, both in French and in positive law." Morbihan MP Paul Molac, the initiator of the law cited, declared: "What absurdity! The only solution to this impasse? Amend the Constitution. Nothing else" (Etienne 2023). Fañch Pichancourt's parents have started preparing their defense for an appeal with the support of *Skoazell Vreizh* and regional councilor Stéphanie Stoll.

In January 2024, another couple, from the Maine-et-Loire (a French department bordering Brittany) were summoned to appear before the family court for having chosen the name *Fañch* with a tilde for their son, born last summer to a Breton mother. Although the civil registrar at the maternity hospital had warned the parents that the spelling of *Fañch* could pose a problem, they had chosen to keep the name, even if it meant having to fight for it if necessary. In this case, according to the Angers public prosecutor, not only is the tilde not a diacritic mark retained by the French language, but moreover the name *Fañch* is considered "contrary to the child's best interests". Loïg Chesnais-Girard, the president of the Regional Council of Brittany, spoke of an "inadmissible social violence" and offered "his full support" to the Maine-et-Loire family, regretting that "for a diacritic sign that is nonetheless not unknown to the French language, as the Roazhon/Rennes Court of Appeal recalled on November 19, 2018, it is time for the law to move forward" (Salliou 2024). Philippe Grosvalet, senator of the south Brittany department of Loire-Atlantique, also intervened with the Minister of Justice, noting: "it is common knowledge that our courts are clogged and that

magistrates are unable to rule within a reasonable timeframe […]. In this context, initiating this procedure is tantamount to adding to the workload of judges who are already far too busy. Worse still, it diverts them from one of their essential missions: ensuring the well-being of the children." The senator concludes his letter with this request: "Put an end to this sterile litigation, which shows little respect for children, parents and the work of family court judges" (Courrier de l'Ouest 2024).

8 A few positive rulings aside from the Fañch issue

While the Martí affair was still underway – a procedural marathon that lasted almost a decade – the *Tribunal de Grande Instance* (ordinary court of first instance) of Perpinyà/Perpignan did validate the civil registration of the Catalan first names *Lluís* (Montpelhièr/Montpellier TGI ruling of September 27, 2001), *Joan-Lluís* (Perpinyà/Perpignan TGI ruling of January 13, 2004) and *Alícia* (Perpinyà/Perpignan TGI ruling of February 16, 2006) in their Catalan spelling, retaining their letter <í> with its acute accent. Conversely however, in 2011, the Tolosa/Toulouse registry office refused to register a little girl named *Alaís* because the letter <í> does not exist in the French alphabet, even though the same office in the same city had previously accepted the registration of a little Anaís and a little Loís (Jornalet 2016) in their correct Occitan forms.

Meanwile, the issue had briefly shifted from anthroponymy to toponymy, as the Occitan spelling of a placename on the Languedoc coast in the South of France had become the subject of headlines. In August 2009, the municipal council of Villeneuve-lès-Maguelone had its historically Occitan name *Vilanòva-de-Magalona* added alongside the existing sign bearing the official French name, marking the town's entrance. In doing so, Vilanòva-de-Magalona was merely following the example of several hundred cities, towns and villages before it, in Brittany, Corsica, the Northern Basque country, Northern Catalonia and other Occitan regions. At the request of the *Mouvement Républicain de Salut Public*, whose name sounds like a deliberate reference to the *Terreur* revolutionary period, the administrative court of Montpelhièr/Montpellier (October 12, 2010) ordered the mayor to remove the signs bearing the Occitan name of his town. This was in ideological continuity with a former decision by Bernard Bonnet, *préfet*[8] of the Pyrénées-Orientales, to bring the municipality of Perpinyà/Perpignan before the administrative court (January 27, 1997) because the elected city council

[8]The State's representative appointed by the central French government in a department or a region.

had asked the local university to translate the city's toponyms into Catalan for the land registry (Duran 2000). The court referred to both the Constitution and the Highway Code, concluding that, on the one hand, "the translation into Occitan is devoid of any historical basis", and that, on the other, the "EB10 and EB20" urban entrance signs, as defined by Article 5 of the decree of November 24, 1967, as amended, include the accented letter <ò> in Vilanòva, whereas "this diacritic sign is not included in any of the appendices to the aforementioned decree, as it does not exist in the French language". The President of the Court concluded that "as a result, the signs in question are detrimental to the clarity of the information required by the obligation of caution and safety imposed on all road users entering a built-up area."

However, the Marseille Administrative Court of Appeal (June 28, 2012) overturned the Montpelhièr/Montpellier ruling (October 12, 2010), rejecting the *Mouvement Républicain de Salut Public*'s claim and ordering it to pay €2,000 under Article L.761-1 of the French Code of Administrative Justice. Indeed, according to the Court of Appeal, "judicial jurisprudence to the effect that the registration of a first name in the civil register must respect the spelling and toponymy of the French language is in no way transposable to spelling rules concerning regional language translations of town entrance signs indicating the French name of a commune". The court considered that the risk represented by the signs in question for the safety of road users has not been demonstrated and invoked the hierarchy of norms. "Articles 4 and 21 of the law of August 4, 1994, regulatory provisions that do not allow for the use of a grave accent on the letter o cannot be applied to the translation into the regional language of a French-language town entrance sign." Thousands of signs in Basque, Breton, Catalan, Corsican, Occitan, etc., often hard-fought for, and obviously carrying the diacritics of each language, are therefore safe, all the more so since Article 8 of the "Molac" law on the promotion of regional languages legalized bilingual signage in 2021. Several municipalities of Lower-Brittany carry a <ñ> in the Breton spelling of their names: Hañveg (Hanvec), Lanleñv (Lanleff), Lañriware (Lanrivoaré), Lañveog (Lanvéoc), Gwiproñvel (Guipronvel), Plougoñ (Plogoff), Ploñger (Ploumoguer), Roskañvel (Roscanvel), Treglañviz (Tréglamus) not to mention dozens of townships, hamlets, localities and streets. Although the actual road signs are now protected, the only authoritative version to be used in any kind of official paperwork remains the French version.

Another ruling concerning the Breton first name *Derc'hen* eventually ended up with a favorable issue too. On January 23, 2018, the registry office of Roazhon/Rennes, following the advice of the Public Prosecutor's Office, had refused to register a child, born in August 2017, whom his parents wished to name *Derc'hen*.

Although the apostrophe is one of the signs authorized in French, the trigram <c'h> – used in Breton ever since the spelling reform of Jesuit Father Julien Maunoir in 1659 – was criticized for its pronunciation ([x], [ɣ] or [h]), which does not exist in French. However, there have been Bretons officially named *Goulc'han* or *Goulc'hen* for a very long time, so the public prosecutor finally reversed his decision concerning *Derc'hen*. Ironically, the names of at least two of the leading protagonists in the Fañch affair, the public prosecutor of Kemper/Quimper, Thierry Lescouarc'h, and the parents' lawyer, Jean-René Kerloc'h, both include <c'h>, which clearly demonstrates the antiquity and banality of this trigram in surnames considered "French". Three days after the local press had reported on the situation, and the mayor of Roazhon/Rennes, Nathalie Appéré, had announced that she was going to ask for the 2014 memo to be amended, the public prosecutor stated in a press release that the memo does not expressly rule on the use of the apostrophe, and since it is a commonly used spelling mark, it can be considered that its use is not formally prohibited. He announced that this issue would be re-examined with the Ministry of Justice in the name of the risk of breach of equality, as the first names Tu'iuvea, N'néné, D'Jessy or N'Gussan had indeed been accepted in Rennes even though the 2014 memo was already in force.

9 "Regional" and "foreign" languages: A common cause

While the interests of so-called "regional" and "foreign" languages may sometimes be pitted against each other in education or the media, a kind of sacred union prevails in the struggle to reclaim the diacritic marks of French citizens. One obvious illustration of this alliance is undoubtedly the story of João Fañch, a baby born in Créteil on August 18, 2019 to a Breton mother and a Franco-Portuguese father. The Créteil town hall having refused him his two tildes, his birth certificate therefore bears Joao Fanch, but his father made a point of expressing his disapproval by crossing out the "chosen first names" box to indicate "imposed first names" instead, and he started leading the legal battle over the tilde with his partner (Delporte 2019).

Alexandra Ibañez was not allowed to pass on to her son Ronin, born in March 2017, her surname in its entirety, i.e. with a <ñ>, alongside Delaunay, the father's name (Serhani 2017). Despite the fact that her own childhood civil status booklet does indeed include the tilde on her surname, the civil registry of the Baiona/Bayonne town hall (department of Pyrénées-Atlantiques, French side of the Basque country), where her child's birth was declared, therefore imposed

the spelling Delaunay-Ibanez on her. According to this great-granddaughter of a Spanish immigrant, this <ñ> is the story of her family, of her origins, and her son thus simply didn't have the same surname as hers. She also thought it was an administrative error when she registered her son at the civil registry office, when the tilde had disappeared from the birth certificate. The town hall in Donibane Lohizune/Saint-Jean-de-Luz, where the family lives, advised her to appeal to the Baiona/Bayonne public prosecutor, who finally signed the rectification on July 29, 2020. Alexandra Ibañez shared the file she compiled on her Facebook page *Touche pas à mon N tildé* which enabled her to win her case after more than three years. Although her name is of Spanish origin, Alexandra Ibañez points out that many Basque first names, such as *Beñat*, *Aña* or *Iñaki*, are affected by the same problem.

One of the active members of the Facebook group is Elodie Billaud Bidegain, who in 2011 faced the refusal of the Baiona/Bayonne town hall to spell her newborn daughter's name *Aña*, a refusal confirmed by the Baiona/Bayonne public prosecutor. Although her daughter continued to be called *Aña* by the family, her official first name remained *Ana*, without the <ñ>, for over a decade. When someone asked for "Ana", she either believed they were addressing someone else, or thought they were mispronouncing it. In January 2022, following the sending of a new file to the Baiona/Bayonne public prosecutor at the end of November 2021, Elodie Billaud Bidegain was finally informed that the latter had granted the change of first name to *Aña* and requested the modification of the act at the town hall. Another young Basque woman called *Aña* officially also got her tilde back in 2023 (Etxezaharreta 2023), 28 years after her birth. There is something ironic in noting that this <ñ>, often brandished in Spain – including in anti-Catalanist demonstrations – as a symbol of *hispanolidad*, has achieved the feat of bringing together under its small sinusoidal banner a large number of objective allies, supporters of the adoption of diacritics present both in the languages of the descendants of immigration (Spanish and Portuguese in particular) and in territorial languages of the French Republic such as Basque and Breton.

10 National representation steps up

The linguists' debate over the first Fañch case, between the public prosecutor and the Court of Appeal in Roazhon/Rennes, confirmed the legal uncertainty faced by parents wishing to give their children names bearing these diacritics. Some courts have ruled that such marks which appear in dictionaries, are not foreign to French and are therefore acceptable. At the time of writing in August 2024, the

Facebook page *Mignoned Fañch* (Fañch's friends) counted 29 officially accepted tildes all over France and there appeared to be no problem any longer in the Basque country whereas elsewhere in France it still seemed to be a real lottery. Parents therefore have no way to be sure whether or not the transcription of a child's first name containing one of these diacritics will be accepted by a registrar, public prosecutor or any other judicial authority. It was in this sense that the Roazhon/Rennes public prosecutor asserted, since the Court of Cassation had not ruled on the merits of the case in its ruling of October 17, 2019, that it was now up to elected representatives to introduce a bill to change the regulation (Zabaleta 2020).[9]

After numerous attempts by various MPs in the past to introduce bills about regional languages, the bill put forward by MP Paul Molac, entitled "*Loi relative à la protection patrimoniale des langues régionales et à leur promotion*", was debated in both houses of parliament. It passed the first reading in the National Assembly but its main provisions were scaled back at that early stage except, ironically, the registration of names spelled in "regional languages" on birth certificates. The French Senate, in session on December 10, 2020, reinstated many of the proposals rejected by the National Assembly, a version which passed by 253 votes to 59. Senator Monique de Marco in her report on behalf of the Committee on Culture, Education and Communication, points out that the Senate adopted an Article in January 2020 to include <ñ> in the list, and stresses that the proposed wording for Article 9 is more protective in that it does not draw up a list but indicates that all regional language diacritic marks are admissible. At the second reading in the French National Assembly, on April 7 and 8, 2021, opponents of the Senate text and the Minister of Education put forward a series of amendments which, if adopted, would have sent the text back for a new "shuttle" between the two Chambers, but their maneuver failed.

On Thursday, April 8, 2021, during the parliamentary day reserved for Paul Molac's small group, *Libertés et Territoires*, bill n°2021-641[10] relative to "*la protection patrimoniale des langues régionales et à leur promotion*" was eventually adopted by a very large majority, including a hundred or so MPs of President Macron's side, against the very advice of the government they supported. Under the Fifth Republic, this is the first law devoted to regional languages to be definitively adopted, with a few specific measures to protect and promote these languages in three areas: heritage, education and public services. Its dual aim is to protect and promote intangible heritage and cultural diversity, of which

[9] See Appendix 2 for a timeline of legislative action on the issue.

[10] https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000043524722

regional languages are one expression. The text recognizes the existence of a linguistic heritage made up of the French language and regional languages, and specifies the role of the State and local authorities in their teaching, dissemination and promotion. It grants the status of "national treasure" to assets of major interest to the knowledge of French and regional languages, such as recordings or ancient manuscripts. Provisions of the "Toubon Law" (August 4, 1994) on the use of the French language now specify that they "do not hinder the use of regional languages and public and private actions in their favor". In the field of regional language teaching, communes with no bilingual schools are now obliged to contribute to the tuition fees of schools offering bilingual teaching in the area (such as Diwan schools in Brittany). The teaching of regional languages is supposed to be generalized as an optional subject within the normal teaching timetable, from kindergarten to high school, based on the model developed in Corsica since the turn of the century. Bilingual signage is now recognized in the law, which clearly authorizes public services to use regional language translations on public buildings, road signs and institutional communications, for example.

This is where, "after the legislative intoxication, the constitutional sobering up comes in" (Arlettaz 2021). On April 22, 2021, 61 petitioning deputies belonging to the parliamentary majority referred the law on the heritage protection of regional languages and their promotion to the Constitutional Council. The article they challenged was Article 6, which the Councilors found nothing to object to, but they took up two other articles on their own initiative, including Article 4 (on immersive teaching) and Article 9 (on diacritic signs in civil status documents). It should be noted that "these two declarations of unconstitutionality are intended to prevent the teaching of a subject which in fact already exists in the public service, and to prohibit the drafting of civil status documents, some of which have been deemed regular by the judicial judge" (Arlettaz 2021).

Where regional languages are spoken, many local and national elected representatives, including from President Macron's party, often voice their protests loudly in the media and on social networks. On May 29, 2021, large-scale demonstrations took place in Gwengamp/Guingamp and Baiona/Bayonne in particular, where in both cases organizers counted up to 10,000 participants. The NGO European Language Equality Network sent a communication to the United Nations Special Rapporteur on Minority Issues to remind the French state of its fundamental constitutional obligations to protect its own linguistic minorities. At the time of publication of the present article, however, no new measures concerning the authorization of diacritic marks have yet emerged.

11 Concluding remarks

On November 1st, 2023, a very strong windstorm called Ciarán severely affected Brittany and other parts of France and Europe. Some Bretons didn't lose their sense of humor however, when they pointed out on social networks that the administrative authorities had widely used a banned diacritic mark on "á" in their official communication about the storm: the Irish acute accent, also known as *fada*[11]. Although anecdotal, this story also shows that France's position is untenable in the long term. Such a typically French issue will certainly not be solved by retreating into strictly technical considerations (i.e., legal and legalistic considerations). In the Martí case, the European Court of Human Rights ruled in favor of France, but its judgment was qualified by the following comments: on the one hand, "given the total absence of consensus between the member states, they had a particularly wide margin of appreciation;" on the other, the government's justification was necessary "for the time being". However, the 21st century justice modernization law (November 18, 2016) contains amended Article 225 which now prohibits discrimination based on the ability to express oneself in a language other than French. The current regulations on diacritic signs are therefore untenable in the short term, as they create a breach of equality based on an illegitimate criterion, which defines discrimination (Blanchet 2017). Indeed, parents who choose first names using languages for which the currently restricted Latin alphabet of French is sufficient have the right to register them; on the other hand, those who choose first names for which this alphabet is not sufficient do not have the right to register the first name of their choice. To take just one last example in French Polynesia – which is on the official UN list of countries to be decolonized – the head of civil status at Papeete town hall explains to Polynesian parents that Tahitian diacritic marks are not allowed (La Dépêche de Tahiti 2015) and cites the apostrophe marking glottal occlusion among these signs, despite the fact that it is perfectly legal, as seen above, unlike the *tārava*, a diacritic which, ironically, is known in French as a... "*macron*."

According to Paul Molac – re-elected MP by a large majority in July 2024 – several solutions remain possible. While amending Article 2 is not the easiest as it would challenge a sacred tenet, a but less ambitious supplement to Article 75-1 could be better adapted. Pragmatically, two sentences could be added about regional languages: "their teaching is determined by law, the immersion method is constitutional. Regional languages are used in public services and civil status" (Molac 2022). The Constitutional Council's decision has re-established the

[11]Which happens to coincidentally mean "fool" in French.

hierarchy of norms, since it takes precedence over a ministerial memo with no normative value. In the case of immersive teaching, the ruling of unconstitutionality was evaded – but with what sustainability? – by the publication of a mere memo (Bulletin Officiel, L'Education Nationale 2021) as it was politically urgent to put out the fire ignited by the announcement that existing immersive teaching was unconstitutional.

It remains to be seen whether the same strategy will apply to the issue of diacritics, which is just one in a much larger issue of linguistic discrimination in France. In terms of language protection, France remains embarrassingly backward, in a position that is increasingly isolating it from democratically comparable states. The self-proclaimed *Patrie des Droits de l'Homme* (Motherland of Human Rights) promotes its own particularism internationally, as well as multilingualism or linguistic diversity within the European Union, yet remains deaf to any similar recognition on its own territory. France strongly opposes the hegemony of English on a global scale, but does what it criticizes English for doing worldwide within its own borders. It is not credible to claim to be preserving cultural diversity in Europe without even subscribing to international commitments recognized by all. It is unseemly to want to protect minorities in other countries while asserting that there are none on one's own territory. France's credibility would be much stronger, and above all, much more coherent if the State were to commit itself to a genuine recognition of its own cultural and linguistic diversity. Let's go back to the original definition of the term "diacritic": this chapter has dealt with it at length in the domain of orthography; but in the medical field, diacritic symptoms serve to characterize a disease, to distinguish it from all others. The single language ideology remains indeed a French disease, and diacritics are but one of the many symptoms that still distinguish France from many other democratic countries in the world, starting with its closest neighbors. Treatments do exist, neighboring countries like Spain and the United Kingdom apply them. It is probably France's turn to take care of its languages, or rather the speakers thereof, and thus take better care of itself.

Appendix A Timeline of the Fañch affair rulings

- *May 11, 2017*: Fañch Bernard born in Kemper/Quimper. The officer initially spells the name Fanch but the first deputy mayor has the tilde added to the birth certificate.
- *July 2017*: Fañch Bernards's parents summoned to appear before the Kemper/Quimper High Court.

- *September 13, 2017*: The Kemper/Quimper High Court reinstates the spelling *Fanch* without the tilde. The parents take their case to the Roazhon/Rennes Court of Appeal, supported by local and regional authorities in Brittany.
- *October 19, 2018*: Roazhon/Rennes Court of Appeal rules against the Kemper/Quimper court and accepts the legality of registering Fañch with its correct Breton spelling.
- *November 22, 2018*: The Roazhon/Rennes public prosecutor's office lodges an appeal in cassation against this ruling. Appeal rejcted on October 17, 2019, on the grounds of a mere formal defect. The Brest public prosecutor summons all the mayors in his jurisdiction to notify him of any new birth of a baby called Fañch.
- *August 18, 2019*: João Fañch is born in Créteil, near Paris, to a Breton mother and a Franco-Portuguese father. The Créteil town hall having refused both tildes.
- *November 18, 2019*: Another baby Fañch born in Montroulez/Morlaix with his name correctly registered. The Brest public prosecutor orders the removal of the tilde. The family starts a legal procedure.
- *April 15, 2020*: Awen Fañch born in Gwengamp/Guingamp. His name is filled in the civil registration form with an <ñ> but the tilde disappears on the birth certificate. The local mayor reinstates the tilde on a new certificate. No reaction so far from the Sant-Brieg/Saint-Brieuc public prosecutor.
- *June 17, 2023*: The mayor of An Oriant/Lorient, signs the birth certificate of yet another little Fañch with a tilde over the <ñ> although the local registry office had initially opposed. In September, the public prosecutor of An Oriant/Lorient has the name modified: Fanch. The parents appeal.
- *July 26, 2023*: Another baby Fañch born in Maine-et-Loire (bordering Brittany). His parents are called to court in Angers on February 15, 2024. According to the president of the Regional Council of Brittany, this is "inadmissible social violence."

Appendix B Timeline of the legislation on the use of diacritics in civil status

- *Thermidor 2, Year II (July 20, 1794)*: A law requires that all public deeds be written in the French language throughout the territory of France.
- *2014*: The Taubira memo states that the only diacritics allowed in civil status records are those authorized by the French language.
- *2019*: The Roazhon/Rennes public prosecutor declares it is up to the elected national representation (National Assembly and Senate) to change the Taubira memo.
- *February 13, 2020*: first reading in the National Assembly of MP Paul Molac's bill on the heritage protection and promotion of regional languages. The main provisions are scaled back, except the registration of names spelled in regional languages on birth certificates.
- *December 10, 2020*: The French Senate reinstates many proposals rejected by the National Assembly and passes a version of the bill by 253 votes to 59. In this version, all regional language diacritics are declared admissible.
- *April 8, 2021*: The Minister of Education and other opponents try to introduce amendments which, if adopted, would shuttle the bill back to the Senate, but their maneuver fails. At the second reading in the National Assembly, Molac's bill is adopted as law n°2021-641 by a large majority, including many MPs of President Macron's side, against their own government's advice.
- *April 22, 2021*: Instigated by the Minister of Education, 61 MPs of the pro-Macron parliamentary majority refer the Molac law to the Constitutional Council.
- *May 21, 2021*: Decision n°2021-818 of the Constitutional Council declares unconstitutional two articles of the Molac law: immersion education in regional languages and the use of regional language diacritics in civil status records.
- *May 29, 2021*: Important language activist and civil rights marches in Brittany and Northern Basque country against this appeal to the Constitutional Council, but to no avail.

References

Arlettaz, Jordane. 2021. *La promotion des langues régionales: après l'ivresse législative, le dégrisement constitutionnel*. Le club de juristes blog. https://blog.leclubdesjuristes.com/la-promotion-des-langues-regionales-apres-livresse-legislative-le-degrisement-constitutionnel/ (5 June, 2025).

Blanchet, Philippe. 2017. *Un ñ breton ou espagnol menace l'unité française*. Le Club de Médiapart blog. https://blogs.mediapart.fr/philippe-blanchet/blog/190917/un-n-breton-ou-espagnol-menace-l-unite-francaise (5 June, 2025).

Bulletin Officiel, L'Education Nationale. 2021. Memo langues et cultures régionales: Cadre applicable et promotion de leur enseignement. *Le Bulletin Officiel de L'Education Nationale de la Jeunesse et des Sports* 47, 16 December, 2021; MENE2136384C. https://www.education.gouv.fr/bo/21/Hebdo47/MENE2136384C.htm (5 June, 2025).

Bulletin Officiel, Ministère de la Justice. 2014. Circulaire du 23 juillet 2014 relative à l'état civil. *Bulletin Officiel du Ministère de la Justice*. page 2. https://www.legifrance.gouv.fr/download/pdf/circ?id=38565 (5 June, 2025).

Conseil Constitutionnel. 2021. *Décision n° 2021-818 DC du 21 mai 2021 sur la loi relative à la protection patrimoniale des langues régionales et à leur promotion*. https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033418805 (5 June, 2025).

Courrier de l'Ouest. 2024. Prénom Fañch: Un sénateur sollicite le garde des sceaux pour faire cesser ce contentieux stérile. *Courrier de l'Ouest* January 12, 2024. https://www.ouest-france.fr/pays-de-la-loire/angers-49000/prenom-fanch-un-senateur-sollicite-le-garde-des-sceaux-pour-faire-cesser-ce-contentieux-sterile-256236dc-b12d-11ee-bd08-22c85604ac0a (5 June, 2025).

Courrier International. 2018. La guerre française contre le tilde. *Courrier International, (after El Periódico de Catalunya)* 10 August, 2018. https://www.courrierinternational.com/article/vu-despagne-la-guerre-francaise-contre-le-tilde (5 June, 2025).

De Bailliencourt, Donatien. 2018. *La reconnaissance des signes diacritiques dans les documents de l'état civil*. https://www.weka.fr/actualite/administration/article_juridique/la-reconnaissance-des-signes-diacritiques-dans-les-documents-de-l-etat-civil-72150/.

Delporte, Fanny. 2019. Val-de-marne: L'administration refuse que le petit João Fañch garde ses accents. *Le Parisien*. https://www.leparisien.fr/val-de-marne-94/val-de-marne-l-administration-refuse-que-le-petit-joao-fanch-garde-ses-accents-05-09-2019-8146446.php (5 June, 2025).

Duran, Marc. 2000. *Bernard Bonnet, un préfet chez les Catalans*. Barcelona: Éditions de la Tempête.

Esteban, Adrià. 2016. En Marti no pot dir-se Martí perquè «a França tot es fa en francès». *El Nacional.* 22 August 2016. https://www.elnacional.cat/ca/politica/marti-nom-franca_108100_102.html (5 June, 2025).

Etienne, Yoann. 2023. *Justice. une nouvelle affaire du tilde de "Fañch" à Lorient?* France Télévisions. https://france3-regions.francetvinfo.fr/bretagne/morbihan/justice-une-nouvelle-affaire-fanch-a-lorient-dans-le-morbihan-2838605.html (5 June, 2025).

Etxezaharreta, Iban. 2023. Ana s'appelle désormais Aña: 28 ans après, le tilde lui est accordé pour son prénom basque. *France Bleu Pays Basque*. https://www.francebleu.fr/infos/insolite/28-ans-apres-le-tilde-lui-accorde-dans-son-prenom-basque-8906673 (5 June, 2025).

European Court of Human Rights. 2012. *Affaire Baylac-Ferrer & Suarez c/France*. https://hudoc.echr.coe.int/app/conversion/pdf/?library=ECHR&id=001-88951&filename=001-88951.pdf (5 June, 2025).

Feltin-Palas, Michel. 2017. *Prénoms bretons: une longue bataille juridique s'annonce*. L'Express. https://www.lexpress.fr/actualite/societe/prenoms-bretons-une-longue-bataille-juridique-s-annonce_1942897.html (5 June, 2025).

Gicquel, Jean-Éric. 2019. Langues régionales et identité culturelle: La bataille des mots en bretagne. In Amané Gogorza & Wanda Mastor (eds.), *Les langues régionales et la construction de l'État en Europe*, 57–70. Paris: LGDJ.

James, Luke. 2023. French government face legal action after banning parents from giving their child a Breton name. *Nation•Cymru* 17 September, 2023. https://nation.cymru/news/french-government-face-legal-action-after-banning-parents-from-giving-their-child-a-breton-name/ (5 June, 2025).

Jornalet. 2016. Se pòt pas apelar "Martí" per çò qu'en França tot se deu far en francés. *Jornalet: Gaseta occitana d'informacions* 24 August, 2016. https://www.jornalet.com/nova/7263/catalonha-nord-se-pot-pas-nomenar-marti-perque-en-franca-se-deu-far-tot-en-frances (5 June, 2025).

La Dépêche de Tahiti. 2015. Vos droits - Le principe de la liberté du choix des prénoms a des limites. *La Dépêche de Tahiti* 23 November, 2015. https://actu.fr/polynesie-francaise/papeete_98735/vos-droits-le-principe-de-la-liberte-du-choix-des-prenoms-a-des-limites_27699447.html. (5 June, 2025).

Le Fur, Laurent. 2020. Le prénom breton de leur enfant refusé par l'état civil, les parents d'Awen Fañch persistent. *L'Écho de l'Armor et de l'Argoat* 14 May 2020. https://actu.fr/bretagne/pabu_22161/le-prenom-breton-leur-enfant-refuse-par-letat-civil-parents-dawen-fanch-persistent_33597360.html (5 June, 2025).

Le Monde. 2019. Le tilde de Fañch définitivement autorisé. *Le Monde* 17 October 2019. https://www.lemonde.fr/societe/article/2019/10/17/le-tilde-de-fanch-definitivement-autorise_6015923_3224.html (5 June, 2025).

Molac, Paul. 2022. Sortir de l'impasse. *Le Peuple Breton* 697. 2.

Ravasi, Elisa. 2017. *Human rights protection by the ECtHR and the ECJ: A comparative analysis in light of the equivalency*. Leiden: Brill.

Rouz, Bernez. 2020. *Fañch; le prénom breton qui fait trembler la république*. Gourin: Éditions des Montagnes Noires.

Salliou, Sylvaine. 2024. Une violence sociale inadmissible: Le prénom Fañch à nouveau devant la justice. *Franceinfo* 9 January 2024. https://france3-regions.francetvinfo.fr/bretagne/le-prenom-fanch-a-nouveau-devant-la-justice-le-president-de-la-region-bretagne-denonce-une-violence-sociale-inadmissible-2902826.html (5 June, 2025).

Serhani, Sophie. 2017. Pays basque: Le combat d'une mère pour donner son nom à son enfant. *Sud-Ouest*. https://www.sudouest.fr/pyrenees-atlantiques/bayonne/pays-basque-le-combat-d-une-mere-pour-donner-son-nom-a-son-enfant-3328736.php (5 June, 2025).

Taubira, Christiane. 2012. Langues et cultures. *Le Club de Médiapart blog* 18 March 2012. https://blogs.mediapart.fr/christiane-taubira/blog/180312/langues-et-cultures (5 June, 2025).

Zabaleta, Eneritz. 2020. *Le ñ de la discorde: retour sur le contentieux Fañch et la transcription des noms et prénoms non français dans l'état civil français*. Revista de Llengua i dret/Journal of Language and Law blog. https://eapc-rld.blog.gencat.cat/2020/01/30/le-n-de-la-discorde-retour-sur-le-contentieux-fanch-et-la-transcription-des-noms-et-prenoms-non-francais-dans-letat-civil-francais-eneritz-zabaleta (5 June, 2025).

Name index

Abbeduto, Leonard, 345, 352
Abels, Klaus, 98, 118–120, 125, 159
Acquaviva, Paolo, 4, 6, 7, 18
Adger, David, 101, 103, 109, 115, 152, 155–160, 170
Aitken, Adam Jack, 254
Alessio, Dominic, 518
Alexiadou, Artemis, 107
Alexopoulou, Theodora, 436
Allen, Liradie, 389
Anand, Pranav, 40, 41
Anderson, Stephen R., 218
ap Dafydd, Myrddin, 232
Ar Bihan, Herve, 439
Arlettaz, Jordane, 546
Arman, Laura, 441
Armostis, Spyros, 394
Armstrong, Timothy, 346
Arnstad, Henrik, 500
Ashby, Michael, 394
Atkinson, Quentin D., 431, 443
Auroux, Sylvain, 417
Awbery, Gwenllian M., 236–238

Baayen, R. Harald, 264, 300
Baker, Jim, 429, 434, 436, 441, 442, 444, 450
Baker, Mark, 120, 424, 427, 441
Baldwin, Timothy, 503
Ball, Martin J., 218, 237, 388, 400
Balles, Irene, 419
Banksira, Degif Petros, 394
Barðdal, Jóhanna, 420, 422
Barker, Chris, 6
Barnes, Michael, 249
Barr, Dale, 300
Barreda, Santiago, 322, 323
Barss, Andrew, 171
Bateman, Nicoleta, 286
Bates, Douglas, 263, 300
Bauer, Michael, 225
Bauer, Winifred, 6
Beck, Sigrid, 162
Beckman, Jill, 386, 388
Beekes, Robert S. P., 418
Beirich, Heidi, 499
Bell, Elise, 236, 238, 388
Bellert, Irena, 106
Benjabi, Ido, 38
Bennett, Ryan, 98, 105, 115, 126, 133, 215, 222, 235, 285, 288, 290, 295, 300, 308, 310, 311, 316, 322, 324, 326, 327
Bergman, Harriet, 496
Berwick, Robert C., 97, 98, 420, 425
Biberauer, Theresa, 118, 426–428
Bidese, Ermenegildo, 67
Bijankhan, Mahmood, 395
Bird, Barbara, 249
Bird, Sonya, 287
Biteeva, Ekaterina, 290, 292
Bjorkman, Bronwyn, 6, 165, 174
Blanchet, Philippe, 536, 547
Blevins, Juliette, 394

Bloomfield, Leonard, 477
Boeckx, Cedric, 425
Boersma, Paul, 262, 297, 322
Bolinger, Dwight, 13
Bolker, Benjamin, 300
Boomgaard, Michelle C., 518
Borgstrøm, Carl H.J., 251, 252, 255, 261
Borsley, Robert D., 67, 70, 188, 189, 191–193, 196, 197, 206, 436, 437, 445, 448
Bortolussi, Luca, 432, 434, 451
Bosch, Anna, 225, 228, 252, 389, 390
Bošković, Željko, 427
Bouckaert, Remco, 431
Bowern, Claire, 422, 452, 453
Brandi, Luciana, 426
Bresnan, Joan, 155
Bridle, Andrew, 497
Broadwell, George Aaron, 158
Broderick, George, 223
Brody, Michael, 120, 121, 132
Browman, Catherine, 325, 326, 328
Brown, Wella, 234
Bruce, Kim B., 42
Bruch, Benjamin, 440
Bruening, Benjamin, 142, 143, 152, 153, 155, 161, 162, 165, 172
Brugmann, Karl, 421
Bumford, Dylan, 46
Bye, Patrik, 124
Byman, Daniel, 496
Byrd, Dani, 321, 328

Cable, Seth, 159
Calderón, Fernando H., 498
Campbell, Lyle, 417, 420, 422, 423
Carlson, Jedidiah, 496
Carnie, Andrew, 99–101, 113, 216, 218, 418
Casey, Louise, 481
Catford, John C., 249
Ceolin, Andrea, 429, 432, 433, 448, 450, 452
Chafe, Wallace, 477
Chasaide, Ailbhe Ní, 226
Chernick, Michael, 318
Cheung, Hintat, 352
Chiblow, Susan, 485
Chierchia, Gennaro, 6, 13, 15, 26
Chitoran, Ioana, 287
Cho, Taehong, 328
Chomsky, Noam, 36, 97, 98, 128, 420, 425, 426
Chung, Sandra, 99
Cinque, Guglielmo, 99, 104–106, 111, 118
Clackson, James, 420, 422, 423
Clark, Robin, 426
Clayton, Ian, 247–252, 255, 261–263, 272, 275, 276, 322
Clemens, Lauren Eby, 101
Clement, R.D., 249, 250
Clements, George Nickerson, 225
Collins, Chris, 142, 163, 167, 171
Condoravdi, Cleo, 15, 26, 47
Cordin, Patrizia, 426
Covington, Michael, 352
Crenshaw, Kimberle, 497
Crisma, Paola, 433, 436
Crnič, Luka, 6
Croft, William, 422
Currie, Oliver, 422
Cyran, Eugeniusz, 386
Czaplicki, Bartłomiej, 287

D'Antuono, Nicola, 11, 30

Daniels, Don, 423, 452, 453
Daniels, Jessie, 496
Dayal, Veneeta, 15
De Bailliencourt, Donatien, 529
De Bhaldraithe, Tomás, 285, 378
De Búrca, Seán, 287
de Gilbert, Ona, 498
de Haan, Germen, 92
de Jong, Kenneth, 225, 228, 252
de Keulenaar, Emilie V., 500
De Melo, Gerard, 380
de Saussure, Ferdinand, 477
de Swart, Henriëtte, 15
Deal, Amy Rose, 5, 6, 16, 176
Declerck, Renaat, 15
Delbrück, Berthold, 418
Delporte, Fanny, 543
den Dikken, Marcel, 32, 33
Deshayes, Albert, 380
Diehl, Randy, 289
Dogil, Grzegorz, 239
Doherty, Cathal, 113
Dorian, Nancy, 252
Doron, Edit, 101, 445
Douglas, Jamie, 428
Downing, Laura J., 397
Dressler, Wolfgang, 418
Dryer, Matthew, 425
Duffield, Nigel, 7
Dunn, Michael, 429
Duran, Marc, 542
Durie, Ross, 417
Dyen, Isadore, 431
Dziadowiec, Ryan, 485

Eckert, Penelope, 275
Edwards, John, 346
Elégoët, Fañch, 88, 89
Elfner, Emily, 109, 129
Emonds, Joseph E., 99, 168, 205
Eska, Joseph F., 212, 283, 391, 395, 398–400, 403, 405, 418, 419, 440
Espinosa, Aina, 327
Esteban, Adrià, 531
Etienne, Yoann, 540
Etxezaharreta, Iban, 544
European Court of Human Rights, 533, 534
Evans, D. Ellis, 418
Evans, Lucy, 489
Eythórsson, Thórhallur, 422

Farkas, Donka, 42
Farnetani, Edda, 327
Fealy, Geraldine, 308
Feltin-Palas, Michel, 534
Ferdinand, Siarl, 230
Ferraresi, Gisella, 418, 419
Ferrer, Joan, 391
Fife, James, 212, 213
Fillmore, Charles J., 423
Flemming, Edward, 328
Flores, Nelson, 497, 511
Folli, Raffaella, 165, 166, 176
Fox, Anthony, 417
Freeze, Ray, 101, 103
Frick, Robert, 308
Friedrich, Paul, 418, 419
Fujimura, Osamu, 288, 291, 320, 321, 324–326
Fullwood, Michelle Alison, 222
Fynes-Clinton, O.M., 198

Gaeloideachas, 346
Gajewski, Jon, 6, 39, 46
Gal, Susan, 497
Garrett, Andrew, 394

Gelman, Andrew, 300
Georgopolous, Carol, 101, 103
Giannakidou, Anastasia, 5, 6, 13, 15, 30, 44
Gick, Bryan, 292, 312, 324–327
Gicquel, Jean-Éric, 537
Gildea, Spike, 419, 422, 423
Gillies, William, 251
Givón, Talmy, 419
Goblirsch, Kurt Gustav, 396
Goldbach, Maria, 418, 419
Goldberg, Adele E., 422
Goldstein, Louis, 321, 325, 326, 328
Grant, William, 251, 252, 275
Graves, Eugene Van Tassel, 438
Gray, Louis H., 391
Gray, Russell D., 431, 443
Green, Antony Dubach, 216, 218, 225, 296
Green, Georgia M., 142, 145
Green, R. Jeffrey, 396
Greenbank, Paul, 473
Greenberg, Joseph H., 98, 419
Greene, David, 284, 397
Gribanova, Vera, 38
Griffen, T.D., 198
Grijzenhout, Janet, 218
Gruzdeva, Ekaterina, 397
Guardiano, Cristina, 425, 430, 433, 437, 452
Guerzoni, Elena, 6, 44
Guilfoyle, Eithne, 100, 101

Haddican, Bill, 167
Haegeman, Liliane, 5, 16, 91, 92
Haider, Hubert, 98
Hale, Kenneth, 147
Hall, Tracy Alan, 385
Hamann, Silke, 397
Hammarström, Harald, 432
Hammond, Michael, 184, 223, 225, 237, 238, 252
Hamp, Eric P., 218
Hannahs, S.J., 198, 205, 206, 218, 236–238
Harley, Heidi, 141–144, 146–149, 152, 153, 165, 166, 176
Harlow, Stephen, 418
Harrell Jr., Frank E, 317
Harris, Alice C., 420, 422, 423
Harrison, Shelly, 421
Hartmann, Frederik, 418
Hartzell, Stephanie L., 498
Haspelmath, Martin, 13
Hay, Jennifer, 155
Heck, Fabian, 159
Heim, Irene, 6
Hejná, Míša, 388
Hemon, Roparz, 230, 231, 445, 518
Henderson, George, 389
Hendrick, Randall, 418
Henry, Victor, 380
Herburger, Elena, 6
Herdan, Simona, 46
Heycock, Caroline, 92
Hickey, Raymond, 216, 252, 285
Hill, Jennifer, 300
Hill, Nathan W., 396
Hind, Kevin, 403
Hinton, Leanne, 496
Hirose, Aki, 394
Hitler, Adolf, 519
Hock, Hans Henrich, 417
Hoeksema, Jack, 6, 46
Holden, Claire, 431
Holmberg, Anders, 92, 167
Holmer, Nils, 252
Homer, Vincent, 6, 13, 39

Honeybone, Patrick, 384, 385, 399
Horn, Laurence R., 13, 16, 26, 50
Hornsby, Michael, 518
Howard, Edwin M., 46
Hsieh, I-ta Chris, 6
Huang, C. T. James, 427
Hueber, Thomas, 297
Humphreys, Humphrey L., 438
Hyman, Larry M., 397

Idsardi, William, 323
Iosad, Pavel, 222, 237, 238, 261, 386–389, 395, 401, 403, 406
Irvine, Judith T., 497
Isaac, Graham R., 399, 400, 419
Iskarous, Khalil, 287, 290, 321, 322
Israel, Michael, 16, 26

Jackendoff, Ray, 106, 107
Jackson, Kenneth Hurlstone, 287, 392, 398, 400
Jacob, Francois, 425
Jäger, Gerhard, 432
James, Luke, 539
Janke, Vikki, 173
Jarvis, Scott, 349
Jaspers, Jürgen, 486, 490
Jeffers, Robert J., 418
Jenner, Henry, 234
Jeong, Sunwoo, 6, 26, 28
Jessen, Michael, 395
Johnson, Kyle, 162, 167
Johnston, Paul, 252
Jones, Charles, 268
Jones, Glyn E., 237, 388
Jones, Mari C, 236
Jones, William, 417
Jongeling, Karel, 429
Jordán Cólera, Carlos, 391
Jørgensen, Anders Richardt, 393, 394
Jornalet, 541
Joseph, Brian D., 406
Jouitteau, Mélanie, 65, 67, 70, 72, 78, 435
Jung, Hyun Kyoung, 142, 146, 147, 165, 171, 176

Kadmon, Nirit, 6, 15, 26, 49
Karttunen, Lauri, 15, 49
Kathol, Andreas, 67
Katsika, Argyro, 312
Kavitskaya, Darya, 287, 290, 321, 322
Kayne, Richard S., 99, 101, 111, 144, 427
Kemper, Susan, 352
Kennedy, Brendan, 498
Kervella, Frañsez, 83
Kibre, Nicholas J., 218
Kikusawa, Ritsuko, 422
King, Gareth, 206, 236
Kingston, John, 289, 291
Kinn, Kari, 428
Kirby, James, 386
Kirchner, Robert, 400
Kirkham, Sam, 287
Kiziak, Tanja, 82
Klatt, Dennis, 322
Klima, Edward, 49
Knuth, Donald, 117
Koch, John T., 399, 400, 418
Kochetov, Alexei, 285–288, 290–292, 296, 316, 326
Koopman, Hilda, 99
Kortlandt, Frederik, 396, 443
Krakow, Rena A., 288, 320, 321, 326
Krasukhin, Konstantin, 418
Krifka, Manfred, 6, 24, 26, 28, 47
Kroch, Anthony, 440

Kroll, Margaret, 40
Kroskrity, Paul V., 496
Kruschke, John, 308
Kuhn, A., 421
Kümmel, Martin, 391, 395, 397, 399
Kuznetsova, Alexandra, 263, 300

Labov, William, 5, 275, 421
Ladd, D. Robert, 386
Ladefoged, Peter, 225, 251, 252, 261, 290, 388
Ladusaw, William A., 5, 6, 39, 49
Lahiri, Utpal, 6, 26
Lainé, Célestin, 518
Laka, Itziar, 5, 6, 30
Lambert, Pierre-Yves, 77
Landman, Fred, 6, 15, 26, 49
Larson, Richard K., 141, 142, 153, 171, 172
Lasnik, Howard, 171
Le Dû, Jean, 232
Le Fur, Laurent, 539
Le Roux, Pierre, 441
Leach, Daniel, 518
Ledgeway, Adam, 425, 428, 437, 438
Lee, Chungmin, 48
Lee, Young-Suk, 26
Lee-Kim, Sang-Im, 320
LeGrand, Jean, 14, 15
Lehmann, Winfred P., 418, 419
Levin, Beth, 142, 145, 147, 161
Lewin, Christopher, 213, 284, 287, 401
Lewis, Henry, 391, 418, 439, 441, 445
Li, Min, 298
Lieber, Rochelle, 218
Lightfoot, David W., 418, 420, 421
Liljencrants, Johan, 289
Lindblom, Björn, 289
Linebarger, Marcia, 47
List, Johann-Mattis, 432
Litvin, Natallia, 287
Llwyd, Alan, 232
Longobardi, Giuseppe, 425, 429, 430, 432–434, 437, 452
Lowe, Jenefer, 213
Lüdecke, Daniel, 303
Lui, Marco, 503
Lynn, Teresa, 353, 354

MacBain, Alexander, 380
MacCana, Proinsias, 418
MacGill-Eain, Tormod, 471–491
Maclean, Norman, *see* MacGill-Eain, Tormod
Macron, President, 536, 545, 546, 550
Maddieson, Ian, 290
Maguire, Warren, 403
Maki, Hideki, 17, 37, 170
Makowski, Dominique, 308
Malmi, Anton, 287
Marantz, Alec, 142
Marotta, Giovanna, 399
Martinet, André, 289, 392, 396
Matasović, Ranko, 380
Mather, James, 249
Matuschek, Hannes, 300
Mayr, Clemens, 44
McCarthy, Philip M., 349
McCloskey, James, 7, 17, 21, 28, 29, 31, 34, 36, 37, 99, 100, 147, 156, 157, 170, 351, 418, 436, 437, 444
McClure, J. Derrick, 249
McCone, Kim, 284, 392, 396–398, 400
McEwan-Fujita, Emily, 517
McMahon, April, 432

McMahon, Robert, 432
McManus, Damian, 393
Medeiros, David, 117, 118
Meelen, Marieke, 419, 425, 432, 435, 440, 449
Meeuwis, Michael, 490
Meid, Wolfgang, 418
Meighan, Paul, 485
Meillet, Antoine, 418
Melchers, Gunnel, 249
Merchant, Jason, 3, 30, 41, 42
Miller, D. Gary, 418, 419
Mittwoch, Anita, 15
Miyagawa, Shigeru, 142
Molac, Paul, 528, 540, 545, 547
Monahan, Philip, 323
Moncunill, Noemí, 391
Mordrel, Olier, 518
Morey, Richard D., 308
Morgan, T.J., 193, 195, 197, 200, 206
Morris, Jonathan, 215, 388
Morris-Jones, John, 232, 236
Morrison, David A., 417
Morrison, Donald A., 132, 403
Mousikou, Petroula, 328
Müller, Nicole, 218, 237
Myers, Scott, 288
Myhill, John, 101

Nam, Hosung, 321
Nance, Claire, 225, 251, 252, 261, 284, 287, 387, 388
Natvig, David, 406
Nearey, Terrance, 323
Neeleman, Ad, 98, 118, 173
Nespor, Marina, 126, 127
Neureiter, Nico, 432, 442, 443
Newmeyer, Frederick J., 425
Newton, Brian, 394
Newton, Glenda Elizabeth, 419
Ní Chasaide, Ailbhe, 251, 261, 285, 308, 387, 388
Ní Chiosáin, Máire, 216–218, 220–222, 285–288, 292, 295, 296, 326, 327
Ní Ghráda, Mairéad, 356
Nichols, Johanna, 429
Nikitina, Tatiana, 155
Niyogi, Partha, 420
Nolan, Brian, 444
Noonan, Maire, 100
Norland, Eric, 86
Nourbakhsh, Mandana, 395
Nurmio, Silva, 449

Ó Baoill, Dónall P., 17, 37, 170, 353
Ó Béarra, Feargal, 346
Ó Broin, Brian, 350, 352
Ó Buaachalla, Breandán, 387
Ó Cadhain, Máirtín, 356
Ó Cinnseala, Colm, 346
Ó Cuinneagáin, Gearóid, 518
Ó Dochartaigh, Cathair, 261, 389
Ó Donnaíle, Caoimhín P., 371, 488
Ó Gadhra, Nollaig, 346
Ó Giollagáin, Conchúr, 249, 476, 488
Ó Maolalaigh, Roibeard, 227, 284, 287, 387, 389, 489, 491
Ó Muirí, Damien, 351
Ó Murchú, Máirtín, 252, 261, 387, 403
Ó Raghallaigh, Brian, 387
Ó Siadhail, Mícheál, 11, 21, 216, 285, 286, 295, 351
O'Rahilly, Thomas F., 387, 403
O'Rourke, Bernadette, 347
Oberpriller, Johannes, 300
Oeberst, Aileen, 500
Oehrle, Richard, 142, 145, 165, 176

Oftedal, Magne, 251
Ohala, John J., 284, 289, 291, 394, 405, 420
Operstein, Natalie, 287
Ormazabal, Javier, 142, 162
Osthoff, Hermann, 421
Ouhalla, Jamal, 437

Pacchiarotti, Sara, 422
Padel, Oliver J., 438, 439
Padgett, Jaye, 285, 287, 288, 290, 292, 309, 314, 326, 327
Panagiotidis, Phoevos, 447
Parrell, Benjamin, 321
Pearson, Matthew, 171, 173
Pedersen, Holger, 391, 393, 418
Pellard, Thomas, 417
Penka, Doris, 5, 20
Perrot, Jean-Marie, 518
Pesetsky, David, 38, 142, 145, 162, 165, 173, 174
Pétur Helgason, 395
Piette, J. R. F., 439, 441
Pinheiro, José, 300
Pires, Acrisio, 421
Polinsky, Maria, 6, 101
Pollard, John, 516
Poppe, Erich, 445
Preminger, Omer, 6
Press, Ian, 230, 436, 439
Proctor, Michael, 320
Puggaard-Rode, Rasmus, 399
Purcell, Edward T., 292
Pyatt, Elizabeth J., 218
Pylkkänen, Liina, 142

Quentel, Gilles, 518

Raimy, Eric, 386
Ram-Prasad, Krishnan, 435
Rama, Taraka, 432
Ramchand, Gillian, 101, 103, 111, 127, 158, 165
Ranero, Rodrigo, 40
Rappaport Hovav, Malka, 142, 145, 147, 161
Ravasi, Elisa, 533
Recasens, Daniel, 327
Reis, Marga, 82
Rezac, Milan, 66, 156
Richards, Norvin, 145, 153
Ringe, Don, 429
Rivero, Maria-Luisa, 72, 77
Roberts, Craige, 41
Roberts, Ian, 111, 192, 420, 425–429, 434, 436, 441, 442, 444, 446–450
Robinson, Christine, 387
Robinson, Mary, 16
Rodríguez, Clara, 327
Roelefsen, Floris, 6, 26, 28
Rogers, Henry, 218
Romero, Juan, 142, 162
Rosa, Jonathan, 497, 511
Rosenberg, Sheldon, 345, 352
Ross, Malcolm, 417, 423
Rouder, Jeffrey N., 308
Roussou, Anna, 420, 426
Rouz, Bernez, 533, 537
Royer, Justin, 176
Rudin, Deniz, 40
Rumsey, Alan, 418
Russell, Paul, 394

Sabban, Annette, 249, 250, 252, 255
Salliou, Sylvaine, 540
Salmons, Joseph C., 386, 388, 406
Sammallahti, Pekka, 397

Sapir, Edward, 477
Sayeed, Ollie, 395
Scatton, Ernest, 287
Schäfer, Florian, 165
Scheer, Tobias, 396
Schleicher, August, 417
Schrijver, Peter, 406, 443
Scobbie, James, 297, 389, 390
Scotland's Census, 476, 477
Ségéral, Philippe, 396
Selkirk, Elisabeth, 126
Serhani, Sophie, 543
Seržant, Ilja A., 423
Sharvit, Yael, 6, 44, 46
Shaw, Jason, 285, 321, 326
Sheehan, Michelle, 428
Shlonsky, Ur, 99, 436
Short, David, 287
Shuken, Cynthia, 248–255, 257, 262, 268, 272, 274
Sidwell, Paul, 395
Silverman, Daniel, 289, 394
Sims-Williams, Patrick, 392, 398, 400, 443
Smith, Jennifer, 433
Smith, Norval, 225, 235, 252
Smitherman, Thomas, 422
Snowdon, David, 349, 351, 352
Snyder, Kieran, 155
Snyder, William, 142
Sommerfelt, Alf, 387
Spaargaren, Magdalena Jeannette, 386
Speitel, H. H., 249, 250
Sportiche, Dominique, 107
Sproat, Richard, 288, 320, 321, 324–326
Stainton, Robert, 42
Steblin-Kamenskij, M. I., 396
Stephens, Janig, 65, 445
Steriade, Donca, 291
Stevens, Kenneth N., 322
Stewart, Thomas W., 218
Stifter, David, 284, 387, 390, 392–394, 396, 401, 402
Stoll, Taja, 287
Stromswold, Karin, 142
Strycharczuk, Patrycja, 328
Stuart-Smith, Jane, 249–252, 255, 261, 268, 275, 388
Stump, Gregory T., 445
Sung, Jae-Hyun, 316
Sutrop, Urmas, 417
Svenonius, Peter, 111, 122, 124, 127
Szabolcsi, Anna, 99
Szemerényi, Oswald, 418

Takagi, Hiroyuki, 161
Takatori, Yuki, 286
Tallerman, Maggie, 189, 191–193, 197, 205
Tannen, Deborah, 477
Taubira, Christiane, 530, 531
Teare, Rob, 234
Ternes, Elmar, 252, 284, 287, 390
Thomas, Alan R., 236
Thomas, Peter Wynn, 400
Thomas, R. J., 370, 380
Thomason, Sarah G., 421, 433
Thoms, Gary, 16, 158, 160, 163, 165, 170
Thráinsson, Hoskuldur, 167
Thurneysen, Rudolf, 396, 418
Toman, Jindrich, 437
Tomaschek, Fabian, 300, 303
Toner, Gregory, 370, 380
Toorians, Lauran, 435
Toosarvandani, Maziar, 40

Trask, R. L, 417
Traugott, Elizabeth Closs, 422, 423
Travis, Lisa, 71
Treves, Alessandro, 433
Trousdale, Graeme, 422, 423
Trudgill, Peter, 250, 275, 276
Turk, Alice, 297, 322
Turton, Danielle, 320

Uhlenbeck, C. Cornelis, 418

Vaillant, André, 418
Valmala, Vidal, 32
van Boxhorn, Marcus Zuerius, 417
Van der Wal, Jenneke, 428
Van Gelderen, Elly, 424
van Sluis, Paulus, 406
Vaux, Bert, 387, 395
Venkatesh, Sudhir, 490
Vicenik, Chad, 395
Villar, Francisco, 400, 418
Vincent, Nigel, 437, 438
Viti, Carlotta, 422
Vogel, Irene, 126, 127
von Fintel, Kai, 6, 46, 49
Von Mengden, Ferdinand, 419, 420

Wackernagel, Jacob, 418
Walkden, George, 417, 418, 420–422, 424–426, 453
Walsh, John, 347
Warner, Anthony R., 439
Watkins, Calvert, 391, 418, 421, 443
Watson, Iarfhlaith, 346
Watson, Seosamh, 390
Webb-Davies, Peredur, 206
Weenink, David, 262, 297, 322
Weir, Andrew, 32
Weiss, Michael, 443, 451
Wells, Gordon, 471, 477, 479, 491
Wells, John C., 249–252, 255, 268, 275
Wentworth, Roy Graham, 389
Wheatley, David, 401
Wheeler, Max, 234
Whitney, Rosemarie, 168
Widdowson, Henry, 476
Williams, Briony, 237, 239, 388, 400
Williams, Nicholas J. A., 388, 439
Willis, David, 88, 189, 419, 420, 422–426, 429, 434, 437, 439–441, 445, 446, 453
Wilson, Andrew F., 496, 516, 517
Withers, Charles, 249
Wmffre, Iwan, 435
Wrench, Alan, 297
Wright, Richard, 289, 291
Wurmbrand, Susi, 65, 81, 90

Yoshida, Masaya, 40

Zabaleta, Eneritz, 545
Zair, Nicholas, 380
Zanuttini, Raffaella, 5, 16
Zec, Draga, 403
Zeijlstra, Hedde, 5, 6, 16, 20
Zenker, Olaf, 517
Zwarts, Frans, 6

Language index

Afrikaans, 450
Amdo Tibetan, 396
Applecross Gaelic, 390
Arabic, 511
Armenian, 420
Aspromonte Calabrese, 438

Basque, 447, 529–531, 542, 544
Berber, 437
Biblical Hebrew, 429
Bosnian, 504
Brazilian Portuguese, 449
Breton (Modern), 63–93, 156, 212, 214, 215, 229–235, 237, 238, 240, 283, 386, 389, 393, 400, 405, 406, 418, 434, 435, 437–442, 444, 446, 518, 527–550
Brittonic, 424, 429, 434, 440
Bulgarian, 287, 385, 439

Catalan, 385, 401, 528–533, 541, 542, 544
Celtiberian, 443
Chaha, 394
Chuj, 176
Classical Arabic, 511
Cornish (Modern), 212–215, 224, 229, 230, 234–235, 237, 238, 240, 418, 434, 435, 438–441, 444–446
Corsican, 529, 542
Czech, 437

Danish, 388, 395, 399, 439, 450
Dutch, 450, 501, 504

East Perthshire Gaelic, 403
Easter Ross Gaelic, 390
English (Modern), 141–143, 149–153, 155, 156, 160–163, 166, 167, 169–173, 176, 177, 294, 295, 320–322, 324, 326, 370, 371, 375, 376, 385, 388, 392, 424, 430, 438–442, 444, 448, 450, 501, 504, 505, 507, 509, 516, 518
Estonian, 501

Farsi, 430, 448, 450
Fiorentino, 426
French (Modern), 324, 376, 378, 379, 385, 389, 394, 438–443, 445, 448–450, 452, 501, 504, 510, 527–550

Gaelic, *see* Scottish Gaelic (Modern)
Gairloch Gaelic, 389
Gaoth Dobhair Irish, 387
Gaulish, 390–392, 401, 434, 443
German, 371, 375, 385, 388, 395, 399, 430, 448, 450, 501, 503, 504, 512
Germanic language family, 425, 434, 438, 442, 447–450, 503
Gothic, 420

Greek, 370, 420, 429, 438, 439, 442, 448, 450
Greko, 438
Guianese Creole, 529, 531

Haitian Creole, 531
Hebrew, 429, 436
Hebrides English, 247–281
Hindi-Urdu, 174
Hittite, 446
Hungarian, 447, 501

Icelandic, 385, 388, 395, 449, 450
Indo-European, 430, 434, 435, 442, 443, 446, 447, 450
Indo-Iranian, 442
Irish (Modern), 3–54, 99, 100, 104, 105, 109, 113, 114, 126, 141–177, 212, 214–224, 226–230, 233–235, 240, 283–296, 299, 308, 310, 316, 320–322, 327, 328, 345–366, 372, 378, 380, 387–389, 391–393, 400, 403, 406, 418, 430, 434–437, 439, 440, 442–446, 448–453, 501, 509, 515–518
Italian, 166, 399, 430, 448–450, 452
Italo-Celtic, 443, 447, 451

Japanese, 394, 447

Kanak languages, 530
Kartvelian, 422
Kerne (Skaer/Banaleg), 73

Latin, 376, 378, 379, 442, 446, 504
Leon (Lesneven/Kerlouan), 75, 77, 79, 80, 87
Leon (Plougerne), 74–76, 79, 81, 82, 84–86, 88, 89, 91, 92
Lepontic, 434, 443

Malagasy, 171, 173
Mandar, 171
Mandarin, 447
Manx Gaelic (Modern), 212–216, 223–224, 226, 227, 234, 240, 284, 287, 406
Middle Breton, 393, 406, 424, 439, 440, 444, 445
Middle Cornish, 406, 424, 435, 440, 444
Middle English, 439, 440, 442
Middle French, 379, 442
Middle Irish, 444
Middle Welsh, 424, 434, 440, 441, 445, 446, 449, 452

Nguni languages, 397
Nivkh, 397, 401
Norse, 380
Norwegian, 450, 501

Occitan, 529, 530, 541, 542
Old Calabrese, 438
Old English, 370, 372, 375, 376, 386, 439, 440
Old French, 376, 378, 379
Old Irish, 99, 284, 370, 376–378, 380, 393, 406, 434, 442, 444, 452
Old Norse, 375, 376
Old Tibetan, 396
Old Welsh, 442, 446

Palestinian Arabic, 436
Phoenician, 429
PIE, *see* Proto-Indo-European
Polish, 450, 452
Polynesian languages, 529, 530
Portuguese, 450, 452, 501, 504

Primitive Irish, 403
Proto-British, 377, 439, 440, 443, 444, 446, 452
Proto-Cariban, 422
Proto-Celtic, 370, 375, 377, 452
Proto-Central-Pacific, 422
Proto-Germanic, 370–372, 376, 377, 512
Proto-Indo-European, 370, 371, 377, 418, 421, 442, 443
Proto-Insular-Celtic, 444, 446
Proto-Italo-Celtic, 443
Proto-Tupí-Guaraní, 422

Romance language family, 434, 438, 443, 447–452, 503
Romanian, 452, 501, 504
Russian, 450, 452

Scottish English, 150
Scottish Gaelic (Modern), 97–135, 141–177, 212, 214, 215, 222, 224–230, 233–235, 238, 240, 247, 249–257, 261, 263, 268, 275, 283, 284, 287, 316, 370, 386–389, 391, 397, 401, 406, 437, 471–491, 515–518
Semitic, 434, 447, 449
Serbo-Croatian, 450, 452
Sinitic, 434
Slavic language family, 370, 434, 448, 450–452, 503
Slovenian, 450, 452, 501
South Argyll Gaelic, 389
Spanish, 443, 448, 450, 452, 501, 504, 510
Standard, 83
Sutherland Gaelic, 389, 390
Swedish, 501
Sámi languages, 397

Tibetan, 396
Treger, 77
Triestino, 430
Turkic, 450
Turkish, 437

Ukrainian, 385
Uralic language family, 450, 503
Urdu, *see* Hindi-Urdu

Vedic, 446
Viceitic, 422

Welsh (Middle Welsh), 444
Welsh (Modern), 183–206, 212, 214, 215, 229, 230, 232, 235–240, 283, 370, 380, 388, 389, 391, 393, 395, 398, 406, 418, 422, 424, 429, 430, 434–437, 439–444, 446, 448–453, 518
Wolof, 447

Subject index

adverb ordering, 111, 112
adversative predicate, 13, 17, 48
AJAX, 377
alternative-evoking, 25–28, 35, 36
aspirating languages, 383–407
ASR, *see* Automatic Speech Recognition
Automatic Speech Recognition, 489

backwards binding, 170–174, 176
BCC, *see* Borer-Chomsky Conjecture
Belfast, 347–349, 366
Borer-Chomsky Conjecture, 424–425, 453
borrowing, 379
Brittany, 528, 533–537, 540–542, 546, 547, 549, 550
broad consonant, *see* velarization
Bunadas, 369–381

causer subjects, 174–177
clefting, 155–158
Clilstore, 479, 488, 489
cognate, 369–372, 380
comparative clause, 45–47
comparative method, 417, 418, 421, 422, 424
conditional clause, 12, 44–45
Consonant Mutation, *see* Initial Consonant Mutation
consonant strength, 383–407
contact varieties, 247–281
contrast, 283–293, 295, 316, 320, 322, 325–329
copular clauses, 97, 101, 105, 113, 125

D-level complexity, *see* syntactic complexity
DCxG, *see* Diachronic Construction Grammar
devoicing, 247, 248, 270–272, 275, 276
diachronic, 372
Diachronic Construction Grammar, 422, 423
diacritics, 528–538, 540, 542–548, 550
dialectal variation, 183, 195, 197, 199–201, 203, 204, 206
dictionary, 369–372, 377–380
Digital Archive of Scottish Gaelic, 489
DIL, 377, 378, 380
direct linearization, 97, 99, 104, 114, 119–121, 124–126, 130, 134
disambiguator, 370, 378
Dublin, 347–349, 366

eclipsis, 216, 218, 219, 224, 226–228, 238, 240
embedding, 71, 73–86
empty categories, 196, 197

endangered languages, 497, 498, 514, 521, 522
epenthesis, 216, 220–222, 224, 228, 230, 235, 237, 238
equative clause, 12
etymology, 369–372, 379–381
European Charter for Regional and Minority Languages, 527
European Court of Human Rights, 533, 534, 547
exceptional exponents, 472, 491

Fañch, 528, 533–541, 543–545, 548–549
filtering, 379
first name, 529–538, 540–545, 547
fragment, 4–6, 19, 20, 30–34, 38, 39, 41, 42
Framework Convention for the Protection of Minorities, 527
France, 527–550
free choice reading, 13, 15, 49, 50
French Constitutional Council, 530, 539, 540

Gaeloideachas, 346
Gaelscoil, 346
Gaeltacht, 345–351, 356–360, 362, 364–366
Gall, 481
genealogy, *see* sloinneadh
gestural coordination, 321
gestural timing, 284, 288–293, 304, 316, 318, 326–328
gesture, 284, 289–293, 299, 303, 304, 307, 314, 315, 320–322, 324, 326–328
GitHub, 377
glottalization, 247, 248, 254, 255, 266–268, 270, 272, 275, 276
Google Translate, 489
GPC, 370, 380
grammatical complexity, 364
Guthan nan Eilean, 471–479, 487–491
Gàidheal, 481, 483, 491
Gàidhealach, 483

Harris, 247, 248, 250, 251, 253–255, 257, 260–265, 267–272, 274
Hebrides, 247, 249–252
heritage languages, 496–498, 508, 509, 521, 522
homograph, 370, 372, 380, 381
hypergraph, 372
hypograph, 372
hypotaxis, 365

implicative predicate, 13, 48, 49
inherently negative expression, 4, 5, 18, 20
inherited, 379
Initial Consonant Mutation, 183–206
inventory, 211, 215, 217, 220, 223, 230, 232, 234, 235, 237, 238
Island Voices, *see* Guthan nan Eilean

JavaScript, 377

language ideologies, 496
language revitalization, 496–498, 516, 522
Laryngeal Realism, 385, 386, 388, 390, 391, 405
lenition, 216, 218–220, 223, 226, 227, 233, 387, 392–396, 399–401, 404

Lewis, 247, 248, 250, 251, 253–255, 257, 259–276
lexicon, 349, 350
light object pronouns, 105, 125, 126
Linear Correspondence Axiom, 99, 104, 105, 116, 128, 134
linguistic mudes, 347, 349
loanwords, 183, 184, 198–201
long head movement, 70
Lower-Brittany, *see* Brittany

MariaDB, 377
minimizer, 8–11, 14, 22, 23, 26–31, 39
minority languages, 498, 514, 522
Mirror theory, 120, 121, 125, 132
morphology, 349–351, 362, 365, 366
multidict, 371
mutation, 213–216, 218, 220, 223, 224, 227, 230, 232–240
mutlidict, 371
MySQL, 377

narrative fronting, 21–38
National Assembly (France), 536, 545, 550
negation, 5–7, 10, 11, 16–18, 20–22, 26–31, 34–37, 40, 41, 49
negative polarity item, 3, 4, 6, 12, 13, 15, 16, 18, 20, 22, 23, 26–34, 36, 38, 39, 41, 44, 49, 50
network database, 369
neutralization (positional), 284, 287
new speakers, 216, 347
nonce loans, 183, 201–202, 204
Norman's World, *see* Saoghal Thormoid
NPI, *see* negative polarity item
Nun Study, 352

object shift, 143, 163, 165–177
Oehrle effect, 142, 143, 151, 163, 165, 166, 173–177
orthography, *see* spelling

palatalization, 215–218, 222, 223, 226–230, 234, 237, 240, 241, 283–331
parameters, 420, 425–434, 436–438, 444, 450–453
Parametric Comparison Method, 425, 429, 431–437, 442–444, 446, 447, 451–453
parataxis, 365
PCM, *see* Parametric Comparison Method
phonetic variation, 383–407
phonetics, 349–351, 362, 366
phonology, 211–241
PHP, 377
pied-piping with inversion, 158–160
polar question, 5, 12, 16, 19, 20, 31, 43
polarity ellipsis, 38
preaspiration, 226, 229, 247, 248, 251, 253–255, 261–264, 266–268, 270, 272, 274–276, 387, 388, 394, 395, 398, 401, 403, 404
prefix, 375, 376
prestopping, 224
probability, 374, 376
Proto-Indo-European, 417
pruning, 375

raciolinguistics, 497
Raidió Fáilte, 347–349
Raidió na Life, 347–349
reanalysis, 420, 423
reflex, 370, 372, 376
regional languages, 528, 530–532, 534–536, 538, 539, 542, 545–547, 550

release, 283, 285, 289–293, 297, 299, 300, 303, 304, 306, 307, 310–321, 326–328
RTÉ Raidió na Gaeltachta, 346

Saoghal Thormoid, 472, 473, 479–489, 491
scope, 3, 7, 10–12, 21, 26, 28–31, 34, 38, 50
second placement phenomena, *see* tense second
Skoazell Vreizh, 539, 540
Skye, 247, 248, 250–255, 257, 259–272, 274–276
slender consonant, *see* palatalization
sloinneadh, 472, 479–480, 486–487, 490
spanning, 121–124
spelling, 377, 378
stack-sorting, 119
stylistic fronting, 64, 65, 70, 73–79, 81–85, 87, 91
suffix, 375, 376
superlative clause, 12
suppletive, 378
surname, 529–531, 533, 535, 537, 543, 544
synchronic, 372
syntactic complexity, 345–346, 351–358, 358–360, 362, 364, 364–366
syntactic reconstruction, 417–467
syntax, 348, 350, 351, 366

T2, *see* tense second
temporal adverbial, 7, 9, 14–16, 28, 47
tense second, 63–93
TG4, 346, 347
tilde, 530, 531, 534–540, 543–545, 548, 549
Toubon Law, 531, 535, 546
tree, 369–376, 379, 380

ultrasound, 293–300, 308–310, 325

V2, *see* verb second
velarization, 217, 226, 227, 283–285, 303, 306, 314, 315, 319, 320, 327, 331
verb second, 63, 65, 78
Villers-Cotterêts ordinance, 535, 537
voicing languages, 383–407
VOS, 99, 101, 103, 104, 114, 132
vowel reduction, 216, 230, 239–241
VP-fronting, 172
VSO, 98, 99, 101–104, 114, 125, 130, 132

wh-island, 17
wh-question, 11, 12, 17, 19, 32, 33, 44, 50
white nationalism, 497–500, 509, 510, 514–517
white supremacism, 495–522
Wikidata, 381
Wiktionary, 369–381
word order typology, 114, 124, 125, 127–131, 135

Foundational approaches to Celtic linguistics

This book showcases the latest research from the world's leading experts on Celtic linguistics. The 15 chapters span a variety of linguistic subdisciplines as well as theoretical and methodological perspectives. Together, these articles highlight critical aspects of contemporary inquiry into the linguistic systems of Breton, Cornish, Irish, Manx, Scottish Gaelic, Welsh and their ancestor languages. The volume is organized around four key sub-areas: (1) syntax and semantics, (2) phonology and phonetics, (3) language Change, historical linguistics, and grammaticalization, and (4) sociolinguistics and language documentation. The volume's papers offer detailed investigations of current theoretical issues in Celtic syntax, semantics, phonology, and phonetics, as well as of language policy and ideology, language weaponization, and diachronic and synchronic language change. These state-of-the-art contributions represent the impressive diversity of the field of Celtic linguistics and emphasize the wide body of work being conducted in the language communities of the six Celtic nations.